Testing conservation applications for environmental DNA techniques through field surveys of presence, richness and assemblage of lotic biodiversity

Harmony C. Patricio, M.P.A.

Australian Rivers Institute and the School of Environment and Science

GRIFFITH UNIVERSITY

Submitted in fulfillment of the requirements of the degree of

Doctor of Philosophy

February 2018

SYNOPSIS

"Each year at the season of the falling of the waters, the people living in the vicinity of the Golden Basin, the home of the Pla Buk [Mekong giant catfish, *Pangasianodon gigas*], join together for the purpose of catching these fish…..The ceremonies connected with the taking of these fish are ancient and have been performed from time immemorial, and carried out once a year."

- F.H. Giles, 1935

This thesis was undertaken to test the potential of the relatively new survey technique of analyzing water samples which contain DNA that has been shed by organisms (environmental DNA) for the study of diversity and distributions of fishes inhabiting tropical and subtropical rivers. The purpose was to gain information on species richness, assemblages, and distributions; with the ultimate aim of improving conservation management of rare and endangered freshwater fishes. The limitations of conventional (observation-based) survey methods to study rare and threatened fishes in large tropical rivers motivated the choice of testing environmental DNA (eDNA) analysis approaches. Specifically, challenges encountered during prior work to conserve the Mekong giant catfish and other rare species fostered the goal of testing eDNA analysis methods. The Mekong giant catfish was once part of a fishery with substantial cultural and socioeconomic import, as described by Giles (1935). However; today the fishery and ceremony surrounding it no longer exists, as the species has become so rare that the capture of wild individuals is banned and it has been assessed as Critically Endangered by the IUCN Red List. The case of the Mekong River and giant catfish represents an archetypical example of the challenges and limitations of using conventional methods to study and conserve rare fishes in large tropical rivers. These rivers host the highest species diversity of any freshwater environment, yet are facing

the highest rates of extinction. This dire situation is partially due to the difficulty of gaining access to data on spatiotemporal distributions that are needed to inform effective conservation actions.

The relatively new technique of using eDNA to access data on species richness, assemblages and distributions may enable access to data from tropical rivers that are not accessible with conventional methods. This thesis compares the benefits and limitations of conventional survey methods with eDNA techniques, within the context of rare species in large rivers. It reviews the development and application of eDNA to study aquatic macroorganisms, specifically fishes. It describes the rapid growth in application of the technique through an analysis of published literature, and finds the majority of studies are focused on developing the methodology. This technique was tested on captive populations of Mekong giant catfish (*Pangasianodon gigas*) in Thailand, and then used to detect wild giant catfish in the Mekong River at locations where it has been captured in the past. The giant catfish was detected in one of six survey locations, and only one sample from the river.

Given the developmental stage of the eDNA technique, the method was further tested in two smaller subtropical rivers with well-known fish species composition. Three different clade-specific (also termed 'universal' or 'generic') fish primers were used to see if they returned congruent results. Clade-specific primers are designed to amplify all species within a given group, such as bony fishes, without amplifying species outside that group, such as other vertebrates. It was found that each primer set had different biases and that eDNA samples did not cluster either by primer set or by location. A subsequent study analyzed whether such variance was related to the choice of sampling location within the river. Given the highly dynamic and heterogeneous

nature of rivers, it is reasonable to assume that eDNA may not be equally distributed along a cross-section within a single hydraulic unit. High variation was found among samples collected at three different points along a cross-section in a pool and riffle in the Brisbane River, Queensland, Australia when using universal primers and eDNA metabarcoding to determine taxa assemblage. Finally, the method was tested on sediment cores collected in a subtropical embayment at the outflow of the Brisbane River to determine whether a chronology of catchment-scale assemblage of freshwater fish communities could be identified. The results were inconclusive, as primarily bacterial sequences were identified from high-throughput sequencing. These results could indicate that eDNA of macroorganisms is not preserved well enough for detection in subtropical sediment cores. This is consistent with the endosymbiont theory that states that mitochondria are descendent from bacteria that have been assimilated into another cell (Sagan, 1967). This has implications for the design of eDNA experiments that target extremely low copy number template, because eDNA studies often target mitochondria rather than nuclear DNA.

The conservation implications of the overall findings of this thesis highlight the fact that all survey methods face limitations when the target organisms are rare aquatic species. Conservation practitioners and resource managers must be creative and forward-thinking in order to gain data on the distribution of threatened species that are needed to inform effective conservation actions. The lack of such data is a contributing factor to the high rates of decline faced by freshwater biodiversity around the globe. The potential of eDNA tools as described in the literature's early stages may have been somewhat overstated, given that the limitations of the method hadn't yet been exhaustively tested. There is still a great deal to learn about how useful this method can

be. The tool seems to work best when using species-specific primers, which amplify a single species, rather than clade-specific primers which can assess total species richness or assemblage through metabarcoding. The method needs further refinement, yet is currently most useful when applied in tandem with conventional sampling methods to survey areas that are difficult to access or for the case of large tropical rivers that have high spatial extent and little knowledge currently available regarding distributions of threatened species.

STATEMENT OF ORIGINALITY

This work has not previously been submitted for a degree or diploma at any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

- Harmony Patricio

TABLE OF CONTENTS

SYNOPSIS I

STATEMENT OF ORIGINALITY V

TABLE OF CONTENTS VI

LIST OF TABLES IX

LIST OF FIGURES XI

ACKNOWLEDGEMENTS XIV

ACKNOWLEDGEMENT OF PUBLISHED PAPER INCLUDED IN THE THESIS XVI

Chapter 1: INTRODUCTION 1

Chapter 2: LITERATURE REVIEW 7

2.1 Summary 7

2.2 Introduction 9

2.3 Evolution and application of eDNA methods 13

2.4 Sampling challenges 18

2.4.1 Detection probability 19

2.4.2 Habitat characteristics and gear selectivity 22

2.4.3 Destructiveness of sampling 23

2.4.4 Taxonomic accuracy 24

2.4.5 Inferring absence 26

2.5 Conclusions 27

Chapter 3: TRAILS OF RIVER MONSTERS: DETECTING CRITICALLY ENDANGERED MEKONG GIANT CATFISH *PANGASIANODON GIGAS* USING ENVIRONMENTAL DNA 31

3.1 Statement of author contributions 31

3.2 Summary 32

3.3 Introduction 33

3.4 Materials and methods 35

3.4.1 Design, in silico and in vitro validation of an assay for Mekong giant catfish 35

3.4.2 In situ validation of Mekong giant catfish assay 36

3.4.3 Selection of sampling sites in the Mekong 38

3.4.4 eDNA sampling in the Mekong 41

3.5 Results 42

3.5.1 In silico, in vitro, and in situ validation of the assay for Mekong giant catfish 42

3.5.2 Field survey results 43

3.6 Discussion 43

3.6.1 Factors influencing detection probability of the giant catfish and recommendations for future eDNA sampling 45

3.6.2 Perspectives for the conservation of the giant catfish and other Mekong fauna 49

Chapter 4: BIAS IN TAXON RICHNESS AND DIVERSITY ESTIMATES WHEN METABARCODING ENVIRONMENTAL DNA FROM RIVER WATER WITH THREE CLADE-SPECIFIC BONY FISH PRIMER SETS 51

4.1 Summary 51

4.2 Introduction 52

4.3 Methods 55

4.3.1 Field Methods 55

4.3.2 Laboratory Methods 56

4.3.3 Bioinformatics Methods 58

4.3.4 Species distribution modeling 59

4.3.5 Data analysis 60

4.4 Results 62

4.5 Discussion 67

4.5.1 Hypothesis 1: Use of multiple universal primer sets will provide more accurate richness estimates 69

4.5.2 Hypothesis 2: Biases occur equally across samples; therefore beta diversity will be consistent among primer sets 71

4.5.3 Conservation Implications 73

4.6 Conclusions 76

Animal Ethics Statement: 77

CHAPTER 5: LOCATION OF WATER SAMPLING POINT WITHIN A RIVER CHANNEL INFLUENCES EDNA METABARCODING RESULTS: THE CHALLENGE OF RELIABLY SAMPLING EDNA FROM RIVERS 78

5.1 Summary 78

5.2 Introduction 80

5.3 Methods 82

5.3.1 Field Sampling 82

5.3.2 Site Survey 82

5.3.3 eDNA Sample Collection 84

5.3.4 Laboratory Methods 85

5.3.5 Bioinformatics Methods 88

5.3.6 Statistical Analysis 88
5.4 Results 89
5.4.1 Velocity 89
5.4.2 eDNA metabarcoding 92
5.5 Discussion 98
CHAPTER 6: eDNA METABARCODING OF SEDIMENT CORES COLLECTED IN MORETON BAY, QUEENSLAND 105
6.1 Summary 105
6.2 Introduction 106
6.2.1 Study area 108
6.3 Methods 110
6.3.1 Sample collection 110
6.3.2 Core stratigraphy 111
6.3.3 Dating of the sediment cores 111
6.3.4 eDNA extraction 111
6.3.5 Primer selection and design 113
6.3.6 Amplification and sequencing 113
6.3.7 Bioinformatics 114
6.4 Results 115
6.5 Discussion 119
CHAPTER 7: GENERAL DISCUSSION AND SYNTHESIS 124
7.1 Next steps for advancing conservation through eDNA tools 130
7.2 Broader implications for conservation 134
8: APPENDICES 136
REFERENCES 140

LIST OF TABLES

Table 3.1: Number of samples with positive detection of Mekong giant catfish from aquaculture ponds in Thailand during In situ validation of species-specific assay. Densities of catfish in each reservoir defined as follows: High: ±12 MGC – wild individuals captured in the Mekong at least 20 years ago ("the parents") (40–50 kg weight), in a pond of 3200 m^2 with a 2–3 m depth; Medium: ±40 large MGC (40–60 kg) – some up to 20 years old + about 100 smaller MGC of less than 10 kg (F1, second generation bred from wild parents), in a pond of 24, 000 m^2 with a 4 m depth; Low: ±10 MGC – 25 cm length (1–2 kg) in a pond of 6400 m^2 with a 2–3 m depth……………………**39**

Table 3.2: Details of sampling sites surveyed during field surveys collecting water samples for Mekong giant catfish eDNA analysis on the Mekong River……………………**41**

Table 4.1: List of primer names, associated sequences, melting temperatures used for study, mean length of target region plus primers when aligned with various reference genomes for fishes from GenBank (NCBI), and references for two primers sets taken from the literature……………………………………………………………………………………………**58**

Table 4.2: The number of OTUs produced and genera identified in rarefied datasets from eDNA metabarcoding with three universal fish markers; summed across all sites, at a 97% clustering threshold. Orders or families that could not be assigned to the genus level are counted once, and scored as a single genus. Counts only include taxa with a proportion of sequence reads ≥5%……………………………………………………………………………**63**

Table 4.3: Mantel test statistics………………………………………………………………………**65**

4.3a. Pearson's correlation coefficients comparing Bray-Curtis indices on raw data, with rarefication at 2500-reads, and excluding OTUs that were unidentified (support <0.5 for matches with GenBank references)

4.3b. Pearson's correlation coefficients comparing Jaccard's indices, with rarefication at 2500-read depth, and excluding OTUs with support <0.5 for matches with GenBank references. The only significant correlation for all Mantel tests was found for comparison of matrices for the Teleo and Ostei12S primers under the above-mentioned filtering protocols (*p = 0.001).

Table 5.1: Velocity of surface water for a pool and downstream riffle at points on transects where eDNA samples were collected. Point A is at the river bank, Point B is 25% of the wetted channel width, and Point C is 50% of the wetted channel width along a straight transect from left bank to right bank. Numbers presented were collected at 1200 hours……………………………………………………………………………………………**92**

Table 5.2: P-values from nested permutational multivariate analysis of variance examining variation in eukaryota and fish taxon richness and assemblage structure, in response to factors Hydraulic Unit (HU), Channel Position (CP) and Sampling Time (T). P-values obtained with 999 permutations; those < 0.05 are underlined……………………**95**

Table 6.1: Sample depth (cm), core stratigraphy, and burial ages of a selection of deposited sediments at site MB2. Sample depths shaded in purple were selected for sequencing of ancient eDNA………………………………………………………………………**112**

Table 6.2: Primer sequences, melting temperatures for PCR………………………………**113**

Table 6.3: List of fish species identified and results from top hit in BLAST runs…….**117**

LIST OF FIGURES

Figure 2.1: Conceptual model of processes of eDNA collection, molecular workflow and analysis with potential sources of type II errors. Some images adapted from Thomsen and Willerslev (2015)……………………………………………………………………………………**9**

Figure 2.2: Results of a literature search of the Web of Science with the search terms "fish" and "environmental DNA" showing the number of publications per year that use eDNA to study fish. Publications are grouped according to their main research topic. In cases where more than one category was addressed, the primary goal of research was used to assign a category. Publications were grouped by three primary categories: survey applications (green, orange, blue solid), method development and evaluation (hatched colors), and desktop reviews (red solid) (see text for further explanation). Search date: 22 February 2018……………………………………………………………………………**16**

Figure 3.1: Sampling sites for Mekong giant catfish in Cambodia, Lao PDR, and Thailand. Details of each sampling site are given in Table 3.2………………………………**40**

Figure 4.1: The location of sampling sites (green circles) within the Mary River catchment, eastern Australia……………………………………………………………………………...**56**

Figure 4.2: Taxon composition from rarefied data (97% clustering threshold) for assigned sequences in each sample based on eDNA metabarcoding with three generic fish markers (Sample ID's indicate site location with underscore to indicate primer name, with _C = Ostei12s, _M = MiFish, and _T = Teleo). Green shading intensity shows relative proportion of total reads (%) assigned to each taxon by each marker. Dendrogram is based on Bray-Curtis distance. Color bar in first row shows marker identity (black = Teleo, red = Ostei12S, blue = MiFish) …………………………………………..**64**

Figure 4.3: Comparison of taxon composition estimates at four sampling sites produced by all three generic primers with eDNA metabarcoding (1-4.x), conventional sampling (1-4.S), and species distribution modeling (1-4.M). Color key indicates probability of occurrence for modeling results, with 100% probability shading for metabarcoding and seining results where taxa IDs were certain, and no color shading where taxa were not detected or observed……………………………………………………………………………………**67**

Figure 5.1: (a) Sampling location (star) within Brisbane River catchment, southeast Queensland, Australia; and (b) Cross-section locations for the pool (yellow line) and riffle (orange line) hydraulic sampling units…………………………..…………………………………**83**

Figure 5.2: Isovel profiles (water velocity contours (cms)) for the pool (a) and riffle (b) cross-sections, showing the point locations for eDNA sample collection (A, B, and C). The triangle indicates the water surface line…………………………………………………..**91**

Figure 5.3: Number of fish OTUs for samples collected at different times, transect points, and hydraulic units, when only OTUs with a read count ≥10 are included, based on metabarcoding with generic fish primers……………………………………………………….**93**

Figure 5.4: Variation in fish assemblages from samples collected at specific times, transect points, and hydraulic units. Data are based on proportion of reads assigned to OTUs within each taxonomic Order of fishes. The total number of OTUs within each Order is shown in parentheses…………………………………………………………………………………**94**

Figure 5.5: Number of eukaryote OTUs for samples collected at different times, transect points, and hydraulic units, (a) for all OTUs and (b) when only OTUs with a read count ≥10 are included, based on metabarcoding with generic eukaryote primers………………**96**

Figure 5.6: Assemblage of Eukaryotes from samples collected within specific times, transect points, and hydraulic units based on proportion of reads assigned to different phyla……….**97**

Figure 5.7: Ordination plots for fish (a. and b.) and for eukaryotes (c. and d.), with labels showing the time of sampling (a. and c.) and channel position location of sampling along transects (b. and d.). Green triangles represent data points for the pool hydraulic unit and blue triangles represent data points for the riffle……………………………………….**98**

Figure 6.1: Map of Moreton Bay, Queensland, Australia with the location where cores were collected indicated by star, and the core sample location ID "MB2". Modified map from Coates-Marnane et al. (2016)……………………………………………………………….....**109**

Figure 6.2: Proportion of reads assigned to different classes. Read counts for each class are shown in parentheses. "None" category did not achieve class-level assignments, and "blank" category received no assignments………………………………………………………....**116**

Figure 6.3: Alignment of the Teleo_F PCR primer (line 11 of alignment) with the top ten hits from a BLAST of the OTU with the highest number of reads. Dots represent nucleotides that are shared with the reference sequence (line 1 of alignment). The three bases at the 3' terminal of the primer do not match the bacterial sequences that amplified (column 15, see box)……………………………………………………………………….**118**

ACKNOWLEDGEMENTS

A great deal of gratitude is owed to those who supported this work and enabled me to produce this thesis. I thank my supervisors, Professor Jon Olley, Professor David Lambert, and Dr. Richard O'Rorke. They provided sustained guidance and encouragement, assisted with design of research protocols, taught me laboratory procedures, funded laboratory costs, and read and edited many drafts of this thesis. Dr. Mark Kennard went above and beyond, and acted as an informal supervisor. Dr. Kennard kindly dedicated a great deal of his time to support field work, data analysis and writing. Dr. Vincent Resh acted as an external supervisor in the beginning of my candidature, and advised on study design for work in the Mekong River. Dr. Simon Linke also advised on study design for the Mekong work and contributed to reviews of Chapter 2. I thank my co-authors and reviewers for their contributions to chapter 3, which was published. Dr. Leon Huynen and Dr. Richard O'Rorke provided extensive training and assistance with laboratory work, and I am extremely grateful to them both for their time and patience. I express my deepest thanks to all those who served as academic advisors and mentors.

I acknowledge the assistance with field work provided by many individuals, and some of which was funded by the World Wildlife Fund – Greater Mekong Program (WWF). For help with field work in the Mekong River, I thank Sinsamout Ounboundisane, staff of WWF, local government officials, and many villagers and fishers who shared local ecological knowledge. For help with field work in Australia I thank Jack Coates-Marnane, Aaron Dunlop, Steven Faggotter, Vanessa Reis, Justin Stout, and Juan Tao. Many thanks also go to Tanzi Smith and the Mary River Catchment Coordinating Committee.

I am deeply grateful to the Australian Rivers Institute, the School of the Environment, and the institution of Griffith University for providing administrative and financial support to enable me to study in Australia. I also thank my family and friends for their consistent encouragement.

ACKNOWLEDGEMENT OF PUBLISHED PAPER INCLUDED IN THE THESIS

Included in this thesis is a paper in *Chapter 3,* which was co-authored with other researchers. My contribution to this paper, of which I was co-first author, is outlined at the front of the relevant chapter. The bibliographic details for this paper, including all authors, are:

Bellemain, E., Patricio, H. C., Gray, T., Guegan, F., Valentini, A., Miaud, C., & T. Dejean. (2016). Trails of river monsters: Detecting critically endangered Mekong giant catfish Pangasianodon gigas using environmental DNA. *Global Ecology and Conservation, 7*, 148-156. doi: http://dx.doi.org/10.1016/j.gecco.2016.06.007

Appropriate acknowledgements of those who contributed to the research but did not qualify as authors are included in the paper and in the statement at the front of the relevant chapter.

Chapter 1: INTRODUCTION

Human demands for water resources have dramatically altered riverine ecosystems worldwide (Vörösmarty et al., 2010). Across the globe, human modifications to the physical structure of freshwater systems, degradation of water quality, invasive species and over exploitation of fisheries have led to freshwater biodiversity being among the most threatened on Earth (Dudgeon et al., 2006, Hermoso et al., 2015). In response to the poor condition of freshwater ecosystems, rehabilitation projects and species conservation have become a priority (Roni et al., 2008, Hermoso et al., 2015). Understanding the spatiotemporal distributions of species is essential for designing effective conservation and restoration strategies. The most commonly applied methods for collecting information on the spatiotemporal distributions of freshwater species are generally labor-intensive, costly, and often constrained or limited in scale and scope; these significant limitations pose challenges for obtaining data useful for management and conservation (Weaver et al., 1993, Lapointe et al., 2006, Baird, 2007, Minamoto et al., 2012). Surveying DNA shed into the environment, termed "eDNA", has been proposed as an alternative, or complement, to conventional survey methods (Pedersen et al., 2015). This thesis aims to test eDNA surveys as an alternative or complement to conventional methods for assessing freshwater biodiversity, especially for rare species in tropical and subtropical river systems.

All living organisms shed DNA into their environment, primarily through release of gametes, sloughing of skin, mucus, scales, or hair, defecation, and urination (Pedersen et al., 2015). In aquatic environments, it is assumed that shed DNA is mixed throughout the water body (Matsui et al., 2001). Sequencing of molecular material extracted from water and soil samples has been used for decades as a tool to study bacteria (Starnbach

et al., 1989, Bruce et al., 1992). The availability of next-generation sequencing technology has made it affordable to combine DNA-based species identification, and markers that amplify multiple species within a group ('clade-specific' or 'generic' primers) en-masse through a process called "DNA metabarcoding" (Taberlet et al., 2012b). Taberlet et al. (2012b) introduced the term as, "high-throughput multispecies (or higher-level taxon) identification using the total and typically degraded DNA extracted from an environmental sample". They note this term also applies to sequencing of bulk samples of whole organisms (i.e. benthic invertebrates). This thesis distinguishes between studies that use metabarcoding of eDNA and those that use samples of whole organisms by using the term "metabarcoding" to refer specifically to high-throughput sequencing of eDNA. Clade-specific primers have been used sporadically since the 1990's for ecological work, such as characterization of metazoan communities from coprolites (Poinar et al., 1998). Further detail on the development and application of DNA metabarcoding is provided in Chapter 2 of this thesis. Recent advances in high-throughput sequencing have driven relatively new use of this tool for the study of macro-organism diversity and distributions in natural water bodies (Ficetola et al., 2008, Thomsen et al., 2012b, Rees et al., 2014).

The application of eDNA methods to inform biodiversity conservation generally is not without its challenges. Some of these include: Estimating detection probabilities given limited knowledge on the effects of environmental conditions on eDNA degradation and transport or various laboratory protocols to extract and amplify eDNA (Deiner et al., 2015, Ficetola et al., 2015); the varied efficiency of universal primers to amplify different taxa (Dickie, 2010); and the need to build up reference libraries of appropriate genome regions with sequences derived from tissue of specimens validated by taxonomists (Pedersen et al., 2015). We might expect that eDNA methods may be

more sensitive than conventional survey methods in certain cases, but their successful application requires careful consideration of detection probabilities.

The ability to successfully achieve true positive detection of rare species and accurate measures of beta diversity is influenced by each step in the eDNA research and analysis process, including: strategy for sample collections in the field, design of primers, choice of DNA extraction methods in the laboratory, amplification conditions, sequencing efficiency and bias, bioinformatic algorithms and thresholds used for assigning taxonomy to sequence data, and existence of reference genes to match with sequences from field results. As this method progresses, it has become clear that environmental conditions and biomass/density of the target species induce high variation in limits to detection (Schmidt et al., 2013, Deiner and Altermatt, 2014, Ficetola et al., 2015). When using clade-specific primers, their design is likely to introduce bias during amplification and sequencing phases (Dickie, 2010). Biased results may also derive from field sampling procedures (Deiner et al., 2015), as the wide range in water depths and velocities present in rivers may influence detections. Sediment cores may provide a stable environment to maintain an eDNA chronology of past biodiversity (Thomsen and Willerslev, 2015, Slon et al., 2017), yet to our knowledge no published studies have tested the potential for river catchment-level biodiversity analysis by examining ancient eDNA in sediment cores at the outlet of a subtropical river.

Overall, few studies examine the capacity of eDNA methods to accurately assess presence and distribution of rare freshwater fishes, to survey total fish richness, measure beta diversity, and account for bias in results based on the design of clade-specific primers or the point location of water sampling. Chapter 2 provides a review of

the relevant literature and contextualizes the use of this molecular tool for conservation applications given limitations on currently available data for rare freshwater species, and challenges of using conventional methods to survey rare species in many environments.

The utility of the eDNA method for determining the presence of rare species is examined in Chapter 3. Water samples from the Mekong River were analyzed to detect the critically endangered Mekong giant catfish *(Pangasianodon gigas*) (Bellemain et al., 2016). Little is known about current population size, but in 2006 a spawner population abundance of 145 was estimated (Lorenzen et al., 2006). They are distributed over at least 1500km of river from Vietnam, and upstream into Cambodia and northern Laos. The only known spawning site is upstream of the first mainstem hydroelectric dam being built in the Lower Mekong River Basin. The rarity of the target species and large water volumes, high depths and rapid velocities in this tropical river make this a challenging test for any survey method. Species-specific primers were used to amplify eDNA extracted from water samples collected in the Mekong River at six locations with historic records of harvest and a high likelihood of the presence of *P. gigas* based on local ecological knowledge.

Chapter 4 describes tests of the eDNA method in the Mary River (Queensland, Australia), where fish biodiversity is previously well-established (Pusey et al., 2004). The aquatic biodiversity of this system has been thoroughly surveyed with conventional methods over the past decade. eDNA sample collection was also conducted simultaneously with seine-netting for this chapter. Clade-specific fish primers were developed and tested to compare resulting richness and beta diversity estimates

established through past and synchronized conventional surveys to those resulting from eDNA metabarcoding.

In Chapter 5 the role that physical variability in river channels may play in causing variance in richness and taxa assemblage estimates based on eDNA metabarcoding with universal fish and eukaryote primers was assessed. Given the high degree of longitudinal and latitudinal variation within river channels, it is plausible that the specific point location of water sample collection could bias eDNA metabarcoding results. Two hydraulic unit types and three positions within cross-sections were tested using surface samples collected in the Brisbane River at sites where species diversity is well known from conventional surveys.

Chapter 6 explores the potential for sediment cores to provide a chronology of species presence and shifts in community composition over long time frames. The literature contains evidence that eDNA molecules adsorb to sediment and may be preserved for long periods within sediment layers (Pietramellara et al., 2007). These findings indicate that sediment samples could provide a more comprehensive source of diversity information for eDNA applications. A well-established body of literature examines pre-historical shifts in species distributions, primarily flora and bacteria, through the examination of ancient eDNA in sediment cores (Willerslev et al., 2003). However, there are few published examples of this approach being applied to aquatic sediments for the study of vertebrates, especially in subtropical environments. To test the validity of the eDNA metabarcoding method for detecting shifts in biodiversity from subtropical aquatic sediments, this chapter included experiments to extract and sequence eDNA from sediment cores collected near the outflow of the Brisbane River in Moreton Bay, Queensland Australia.

Chapter 7 provides a concluding summary of the experiments performed and the most relevant findings produced, within the context of establishing more effective tools for the survey, management, and conservation of rare and threatened freshwater species. The primary goal of this work was to apply relatively new eDNA techniques to environments and species that are underrepresented in published literature, in order to determine their utility and test their potential for enhancing the sensitivity of diversity surveys. This chapter also highlights the need to explain the benefits, limitations, and relevance of results from eDNA research to conservation practitioners, resource managers, and policy makers. Such knowledge transfers will enable the tool to mature from a developed-nation research enterprise, to a system supporting the global expansion of diversity knowledge and effective conservation of freshwater biota in the regions where it is most threatened.

Chapter 2: LITERATURE REVIEW

2.1 Summary

The global extinction crisis is accelerating and freshwater species are declining at the highest rates. Our success or failure in conserving these species rests heavily on knowledge of their distributions, yet these data are typically non-existent, outdated, or plagued with gaps. Data are especially limited for rare species from tropical environments. This is primarily because conventional observation-based survey methods have low detection probabilities for rare aquatic organisms. Relatively new molecular techniques may facilitate collection of the reliable distribution data needed to conserve the planet's most threatened species; yet there are currently significant difficulties associated with the application of these techniques. This review describes the challenges associated with applying conventional methods to study the distribution of rare aquatic species, describes how molecular techniques may help to overcome them, and clarifies aspects of molecular techniques which need further development to facilitate wider conservation applications.

There are significant data gaps regarding the distribution, composition, and richness of species in many freshwater environments, partially due to the limitations of conventional methods for detecting rare aquatic species, especially in tropical river basins where species richness is highest. Molecular methods may help to overcome these challenges and improve our understanding of rare species distributions, yet further development and validation is required for the broader application of these methods. Analysis of environmental DNA (eDNA) shed by organisms and collected with water samples may provide increased sensitivity for detection of rare species compared to conventional methods in certain cases. This approach has been effective in temperate

regions with moderate species diversity, but requires further refinement before it can be mainstreamed into standard sampling programs. This refinement is especially relevant for rare taxa in tropical or subtropical systems with high richness. Errors in detection can occur at many stages of the eDNA workflow, from sample collection, laboratory protocols, and data analysis (Figure 2.1). Given these challenges, eDNA methods require further refinement and the published literature aimed at improving the technique is rapidly growing (Figure 2.2).

Overcoming the limitations of eDNA and conventional methods, most likely through complementary application of both approaches, could provide essential data for regions with high conservation priority. Calculating detection probabilities in advance and interpreting results with site occupancy modeling probabilities (Schmidt et al., 2013) are essential for useful application of these data. This currently represents a growing research area (Moyer et al., 2014, Ficetola et al., 2015, Schultz and Lance, 2015, Furlan et al., 2016, Schmelzle and Kinziger, 2016, Willoughby et al., 2016). The physical and chemical characteristics of the sampling site and the target organism have substantial influence on detection probabilities. This is because such factors control the amount of eDNA that is released into the environment, how long it persists after release, and how much is available for collection in a sample at any given location or time (Barnes and Turner, 2015). Laboratory methods influence detection by controlling the quality and type of DNA sequences that become available for analysis (Thomsen and Willerslev, 2015). Bioinformatic analysis algorithms and content of reference genome libraries will influence which sequences are retained for analysis and the outcomes of taxonomic assignment (Evans et al., 2017a). (Figure 2.1).

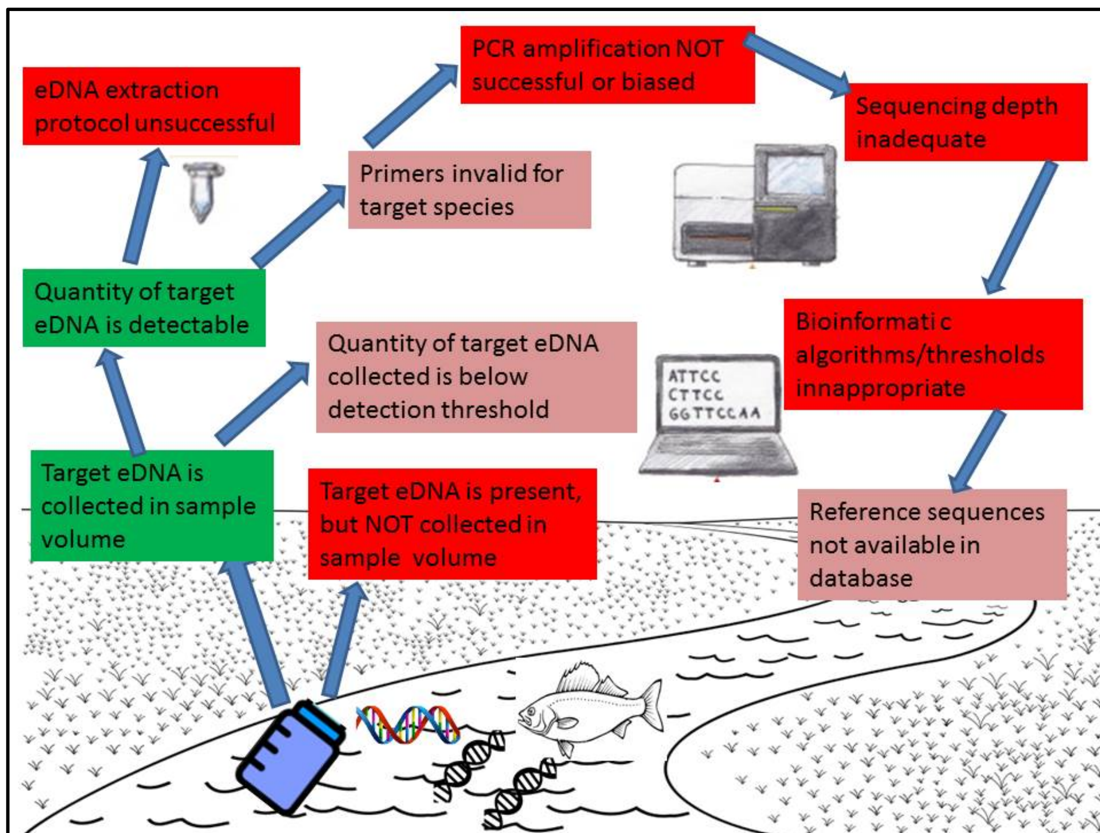

Figure 2.1: Conceptual model of processes of eDNA collection, molecular workflow and analysis with potential sources of type II errors. Some images adapted from Thomsen and Willerslev (2015).

2.2 Introduction

Amidst a global extinction crisis (Barnosky et al., 2011), freshwater species are declining at higher rates than either terrestrial or marine taxa (Ricciardi and Rasmussen, 1999). One principal factor in their decline is the high level of human demand for freshwater resources. Another, often overlooked, factor is that distribution data are not available at the appropriate spatiotemporal scales to enable relevant conservation actions. In order to conserve rare freshwater species effectively, a great deal more distribution and species richness data at finer spatial and temporal resolutions are needed. Species distribution models can be informative, but improving data quality for rare species is a high priority for developing more effective and efficient models (Hermoso et al., 2013). In this review, 'rarity' is considered as those species

with small areas of occupancy and/or which occur in very low abundances (Flather and Sieg, 2007). Rare species occur naturally, but also often include those that are threatened by human activities (i.e. have undergone severe declines in distribution and abundance). Rare species and the locations in which they occur are thus often the focus of conservation management efforts to promote species persistence.

Although some freshwater systems have been sampled extensively, data needed to accurately assess distributions are typically lacking for rare species. The amount of time and effort required for detecting rare species in large waterbodies, and the necessity to minimize harm to the target organisms, often makes such research using potentially destructive sampling methods impractical. The lack of data on rare species distributions is linked to the fact that conventional methods for sampling aquatic organisms are often not well-suited to detecting rare species. This is because these methods require direct capture or observation of individuals which may be harmful, they suffer from low detection probabilities, and they cannot be applied uniformly across different habitat types (Olson et al., 2012). These limitations have driven the search for new sampling approaches that are accurate, sensitive and cost-effective in determining the occurrence of rare species. One emerging approach involves molecular methods that allow the detection of species through the collection and sequencing of small degraded fragments of cellular or extracellular DNA that are shed by living organisms into their environment, often through hair, scales, skin cells, mucus, gametes, or excrement (termed "environmental DNA" or "eDNA"). This survey technique has the potential to provide cost-effective species distribution and richness data to complement labor-intensive and potentially invasive conventional sampling methods (Evans et al., 2017b). eDNA can also be applied in habitat types that are typically inaccessible or dangerous to access with conventional sampling methods, such as large rapids, deep

pools, polluted areas, or locations with dangerous organisms (Evans et al., 2017b). However, this technique cannot provide information on microhabitat use, or distinguish between different sexes, life stages or dead/live individuals. It requires substantial background work to develop appropriate primers and DNA sequence reference libraries in order to identify taxonomy (Roussel et al., 2015, Evans et al., 2017b). Despite the appeal of eDNA methods to rapidly survey diversity and occupancy, they should not be considered a replacement for conventional survey methods. All survey methods have limitations, and the context of survey and research goals should be used to inform the choice of whether to apply eDNA methods or to use them as a complement to conventional survey techniques.

There are a large number of variables to consider when designing an eDNA sampling strategy. Lotic systems may have lower and more variable detection probabilities in comparison to small or medium lentic environments based on factors like fluctuating discharge rates, variation in substrate type, and heterogeneous channel morphology (Civade et al., 2016, Shogren et al., 2017). The persistence of eDNA once it is shed is highly influenced by environmental factors such as temperature, UV light exposure, pH, dissolved oxygen, and levels of microbial digestion of DNA molecules (Moyer et al., 2014, Eichmiller et al., 2016a). The respiration rates, behavior, abundance, density, and biomass of target organisms are a few of the variables that will influence detection sensitivity for different organisms (Darling and Mahon, 2011). One study showed that transport distance downstream and detection limits of eDNA for two macroinvertebrates varied widely in a temperate river, showing that different species likely have different detection probabilities (Deiner and Altermatt, 2014). There is work indicating that eDNA concentrations are correlated with biomass (Kelly et al., 2014),

potentially enabling abundance estimates. However, there does not appear to be a consistent pattern in the relationship amongst different species (Klymus et al., 2015).

Every step in the field sampling and analysis process influences detection rates and accuracy (Figure 2.1). The volume of water collected, the method of collection, and stringent contamination controls are known to be highly important (Darling and Mahon, 2011, Schultz and Lance, 2015, Goldberg et al., 2016). The entire molecular workflow from DNA extraction, to amplification, to DNA sequencing, also has a substantial role in the quality of results. The specificity of primers must be confirmed *in silico* and *in vitro* and assays must be validated, especially in the case of sympatric, closely related, or hybrid species (Wilcox et al., 2013, Fukumoto et al., 2015). The methods employed for storage, extraction, amplification, and sequencing of DNA all have significant effects on the final results of an eDNA study, and there is currently not a consensus on best practices beyond those designed to eliminate contamination. Ficetola et al. (2015) have demonstrated that a wide variety of methods are currently in use, with each having potential limitations.

This literature review evaluates the potential for recent advances in eDNA recovery and next-generation sequencing to help overcome the limitations often presented by conventional methods when sampling for rare aquatic organisms. By combining DNA-based species identification, next-generation sequencing, and universal or clade specific primers that amplify multiple species en-masse, 'metabarcoding' is enabling rapid biodiversity assessments (Taberlet et al., 2012b, Ji et al., 2013). These techniques have the potential to greatly expand the collection of data on rare species distributions, yet still face multiple challenges to widespread application in a conservation context.

This review covers the development of eDNA techniques through time, and assesses trends in the literature focused on fish since the first publications in 2011. Some of the limitations of conventional methods are addressed and the potential for eDNA to help overcome these challenges is described, while noting the current limitations associated with eDNA methods. Recommendations for future directions in research and the importance of providing interpretation of eDNA results in a format that is relevant for conservation practitioners and resource managers are highlighted.

2.3 Evolution and application of eDNA methods

Living organisms continuously shed DNA molecules into the environment. For aquatic species, DNA is shed directly into the surrounding water. By extracting and sequencing this shed eDNA from a water sample it is possible to determine which species have been present near the sampling location. The technology that enables this method is not new, but recent developments in next-generation sequencing have made it feasible to obtain large amounts of sequence data at a relatively low cost (Thomsen and Willerslev, 2015). eDNA can be analyzed with methods such as end-point PCR and quantitative PCR (with Sanger sequencing) to gather species-specific information, and community-level assemblage data can be accessed through metabarcoding with next-generation sequencing. Primers designed to amplify a single species are used for end-point and quantitative PCR with Sanger sequencing, and primers designed to amplify many species within a group are used for metabarcoding.

The methods used in eDNA extraction and analysis rely on the foundational work of ancient DNA (Higuchi et al., 1984, Pääbo et al., 1989, Willerslev et al., 2003) and microbial ecology (Ogram et al., 1987). Many of the methodological and practical challenges of detecting ancient DNA and interpreting results from environmental and

macrofossil samples are similar to those for eDNA, because both typically contain small amounts of highly degraded target DNA. One of the first examples of the application of eDNA in aquatic environments was conducted over 25 years ago, when DNA of whales was extracted from sloughed skin recovered from water samples (Amos et al., 1992). More than thirteen years ago, the first application of eDNA techniques to detect macroorganism DNA in an aquatic system used river water to amplify cow, human, pig, and sheep DNA (Martellini et al., 2005). Some of the earliest work using eDNA to study aquatic species showed that this approach was more sensitive than conventional (observation-based) methods for detecting invasive bullfrogs (Ficetola et al., 2008). One of the first papers using eDNA for freshwater fish demonstrated the ability to detect invasive carp species in very low abundance and in locations where conventional methods failed to detect them (Jerde et al., 2011).

Metabarcoding has been applied to eDNA studies for fish and a range of other aquatic taxa. One of the first successful attempts at metabarcoding aquatic vertebrates using DNA from environmental samples was published in 2012, where the eDNA of four different fish species was detected from river water samples collected in Japan (Minamoto et al., 2012). This study, however, used cloning and Sanger sequencing rather than next-generation sequencing, and therefore had less coverage than what can be done with next-generation sequencing. Rapid uptake of these methods is evidence of their utility, but there is a need for further testing and refinement to improve the usefulness of eDNA methods for conservation and management applications.

The results of a systematic literature search show that eDNA research on fish is a growing field with a rapid increase in the number of publications since 2011 (Figure 2.2). An ISI search of the terms “fish” + “environmental DNA” showed that despite the fact that the number of publications has increased rapidly in recent years, relatively few

focus on the wider application of the technique for conservation and management. The majority of papers are focused on method development, demonstrating that this approach is still being validated to ensure the method provides accurate results for different species and environmental conditions.

The systematic review found that 75 publications focused on methodological refinements to evaluate shortcomings and improve the application of eDNA for fish (Figure 2.2). These methods papers are separated into four primary categories: 1) Sequencing for building reference libraries; 2) Primer design and validation; 3) Protocols for field collection, molecular workflows, and analysis; and, 4) eDNA detection. For example, 27 papers were identified that focused on evaluating factors influencing eDNA detection ("Methods – DNA detection"). These included experimental and applied laboratory and field studies to quantify environmental and biological factors that influence eDNA shedding rates, persistence/degradation rates, transport, retention, and detection probabilities. A total of 29 papers focused on design and assessment of suitable protocols for field sampling, molecular workflows in the laboratory, and data analysis ("Methods – Protocols"). A smaller number of publications focused on methods for primer design (14 papers) and sequencing for reference databases (5 papers). The field of eDNA analysis has also seen a consistent number of reviews and comments through time, and 20 such papers were returned in the ISI search. While some early reviews tended to overstate the potential of the method, recent years have seen reviews that more fully describe the limitations of eDNA and argue for a set of standardized guidelines (Goldberg et al., 2016).The latter includes

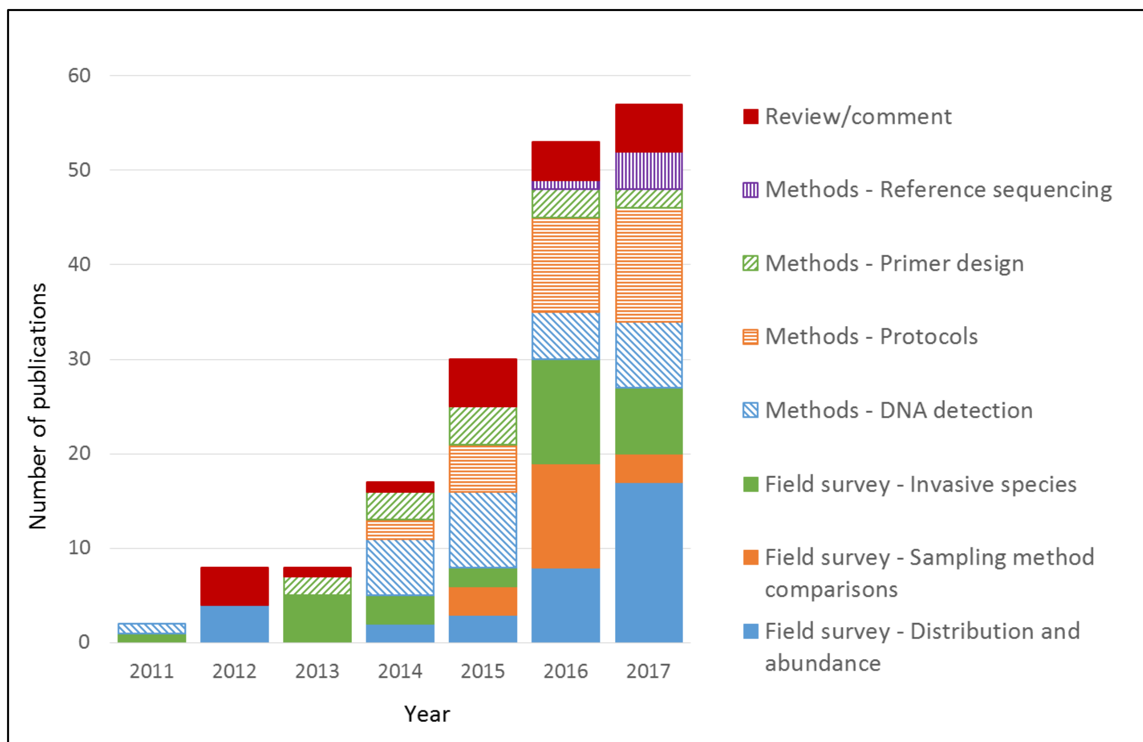

Figure 2.2: Results of a literature search of the Web of Science with the search terms "fish" and "environmental DNA" showing the number of publications per year that use eDNA to study fish. Publications are grouped according to their main research topic. In cases where more than one category was addressed, the primary goal of research was used to assign a category. Publications were grouped by three primary categories: survey applications (green, orange, blue solid), method development and evaluation (hatched colors), and desktop reviews (red solid) (see text for further explanation). Search date: 22 February 2018.

increased field sample replicates, increased PCR replicates, rigorous tests of primer specificity, and use of occupancy models for determining accurate detections and knowledge of detection probabilities (Willoughby et al., 2016).

For publications that applied eDNA methods for fish, the most common (33 papers) reported field surveys of fish distribution/abundance. These included detecting rare species distributions, estimating species' abundance or biomass, and/or quantifying fish community composition and biodiversity attributes such as species richness ("Field survey - Distribution and abundance" in Figure 2.2). A number of field studies (17 publications) have also compared the performance of eDNA methods with

conventional sampling methods to estimate species distributions and/or abundance ("Field survey - Sampling method comparisons" in Figure 1). These studies used either simultaneous sampling or data that were gathered previously with conventional tools such as seine netting or electrofishing. eDNA has also been widely used (31 publications) to detect invasive species occurrence, spawning and spread ("Field survey - Invasive species" in Figure 1), especially for invasive Asian carp in the Great Lakes region of the USA (i.e. Jerde et al. (2013)).

The rapid increase in fish eDNA publication numbers highlights the potential of this approach for a range of conservation and management applications, and demonstrates the need for alternative tools to increase detection and decrease effort for surveys of fish diversity and distributions, especially for rare species. However, the large number of publications focused on method refinement and validation indicates that further work is needed for this approach to be applied in a broad conservation context. Sampling fish with conventional methods often presents challenges, and the following section (2.5) reviews these challenges along with the potential for eDNA approaches to help overcome them. This section also describes the limitations that currently exist for eDNA techniques to address these challenges. Given that the focus of this thesis is to test the utility of eDNA methods for broader application to conservation, the fact that large tropical rivers host the highest number of threatened species and are the most under-surveyed is an important consideration for this review. eDNA methods have primarily been applied for fish research in temperate environments with moderate species diversity and environmental conditions more favorable for eDNA persistence. However, some research has been done in tropical regions of Australia (Robson et al., 2016) and Southeast Asia (Bellemain et al., 2016). Alternatively, rare species and biodiversity hotspots are present in higher proportions in tropical regions. Large

tropical rivers represent one of the most challenging environments to sample with any survey method, either conventional or molecular, because of the high variability of environmental conditions and high water volumes.

To improve distribution and biodiversity data, we need to apply alternative methods, such as metabarcoding of environmental samples, which take advantage of current technologies. Advances in next-generation sequencing have made this tool accessible and affordable (Valentini et al., 2009, Evans et al., 2017b); with especially high utility for rare species sampling. Accessibility of rapidly expanding reference sequence databases, like those provided by the International Nucleotide Sequence Database Collaboration (Benson et al., 2012) and the Consortium for the Barcode of Life (Ratnasingham and Hebert, 2007), reduces the effort and expense of sequencing these reference materials. Where species are rare and distribution and biodiversity data are limited, eDNA methods could provide essential information to enable development of more accurate models and more effective conservation interventions. Designing and testing primers that can accurately amplify and facilitate the sequencing of a large number of species simultaneously through metabarcoding should be a high priority for advancing the conservation potential of this technique.

2.4 Sampling challenges

There are several issues that limit the ability to collect distribution data for rare aquatic species through both conventional and molecular methods. The primary factors for conventional methods include; 1) detection probability, 2) sampling gear limitations, 3) destructiveness, 4) taxonomic accuracy, and 5) failure to infer absence. It is possible that eDNA methods can overcome these challenges in certain cases, yet they face their own limitations. The primary limitations of eDNA methods include 1) ability to quantify detection probabilities, 2) uncertainties for sampling volume

requirements in large rivers, 3) biases from sequencing influenced by universal primer design, 4) uncertainties about the influence of environmental variables such as velocity and hydraulic unit type on collection of eDNA in the field, 5) uncertainties about transport and degradation rates, 6) inability to distinguish sexes, life stages or dead/living individuals, 7) limited availability of reference sequences in online databases, 8) false negatives, 9) contamination, 10) taxonomic resolution of the small region of DNA between primer sites. The following chapters in this thesis address some of the challenges from the list above (numbers 2, 3, 4, and 5) in an effort to advance the conservation utility of the method.

2.4.1 Detection probability

The probability that species or their eDNA are detected given presence depends on many factors, including the susceptibility to capture through chosen sampling gears, the method of sampling selected (i.e. frequency, water volume, spatial extent) and the scale of the survey. Targets may be widespread, but present in low abundance at sites of occurrence, or they may be present in very few sites at higher abundance. Either type of distribution may be exhibited by rare species, and will influence the probability of detection. Behavior of target organisms and their vulnerability to collection gears (nets, traps, etc.), the sensitivity of detection tools, and the way sampling surveys are designed, are determinant factors in successful detection. The probability of detection for rare species is typically low (Bayley and Peterson, 2001), type II (false negative) errors are common, and substantial effort is often required for accurate detection with conventional methods (Olson et al., 2012, Sigsgaard et al., 2015, Evans et al., 2017b). These challenges have been a motivating factor in the rapid uptake of molecular methods. Some studies which sampled simultaneously with conventional and molecular methods have shown that eDNA analysis can sometimes be more sensitive

for detecting rare species, under certain conditions and with particular molecular workflows (Rees et al., 2014, Sigsgaard et al., 2015, Schmelzle and Kinziger, 2016).

Survey gaps from low detection can in part be alleviated by species distribution models that rely on relationships between environmental and physical indicators. These models rely on variables such as catchment topography, climate and hydrology (Scheller et al., 1999, Rose et al., 2016). The use of surrogate environmental and physical indicators to model distributions alleviates the bias associated with survey data gaps, but may introduce the problem of false presences and false absences, which can lead to over/underestimation of occurrence (Loiselle et al., 2003). Models that are biased towards false presence may result in subsequent conservation interventions not actually benefitting target species while consuming precious conservation dollars (Wilson et al., 2005). Models that are biased towards false absence may miss the opportunity to protect a remnant population important for species persistence. In general, the inclusion of rare species in models is especially problematic (Hermoso et al., 2013). Predicting species distributions based on environmental factors requires comparative studies that demonstrate correlations between abiotic factors and species presence (Olden and Jackson, 2002); however, the data needed to examine such correlations are limited for most rare aquatic species and tropical rivers. If applied rigorously, eDNA has the potential to validate modelled distributions.

eDNA benefits and limitations

Application of eDNA technology may increase detection probability over that of conventional methods for some species (Schmelzle and Kinziger, 2016, Shaw et al., 2016, Doi et al., 2017), but not for others (Shaw et al., 2016, Ulibarri et al., 2017). Particular studies showed that eDNA methods can provide more accurate data on presence or higher richness counts when compared with direct observational methods

(Jerde et al., 2011, Dejean et al., 2012, Schmelzle and Kinziger, 2016, Shaw et al., 2016, Doi et al., 2017). While other studies showed that at times conventional methods provided detection when eDNA did not (Valentini et al., 2016, Perez et al., 2017). When preparing for an eDNA field survey, it is important to first test limits of detection through the use of tools such as control experiments and serial dilutions so results can be interpreted appropriately (Ficetola et al., 2015, Hunter et al., 2017). Control experiments could consist of water collection from tanks with varied density of animals, to determine DNA shedding rates. Serial dilutions use a constant dilution factor for a stepwise dilution of DNA, and can determine at what concentrations eDNA must be present in order to be detectable in samples. Some work suggests that sampling strategies, laboratory protocols, and data analysis methods that enable detection probabilities of at least 0.95 should be identified prior to field sampling (Schmelzle and Kinziger, 2016). These publications encourage the use of site occupancy modeling (Schmelzle and Kinziger, 2016), testing of detection limits (Hunter et al., 2017), and testing of required number of PCR replicates (Ficetola et al., 2015) prior to field campaigns.

Presence/absence detections are always susceptible to error, and it is never possible to demonstrate absence with 100% confidence. One new horizon in eDNA research is in understanding detection probabilities to allow the accurate interpretation of negative results. In the case of highly threatened species, a false-absence determination presents risk for management, as it may preclude conservation action in an isolated or remnant population which is the last holdout of a disappearing species. A body of work is currently developing, with useful guidelines for determining detection probabilities, as well as sampling volumes, replication levels, and model parameters needed to reduce the probability of false negative detections and increase

the accuracy, sensitivity, and precision of data collected through eDNA methods (Pilliod et al., 2013, Díaz-Ferguson and Moyer, 2014, Moyer et al., 2014, Deiner et al., 2015, Ficetola et al., 2015, Jane et al., 2015, Schultz and Lance, 2015, Goldberg et al., 2016, Mächler et al., 2016, Wilcox et al., 2016). The following chapters of this thesis contribute to these advances.

2.4.2 Habitat characteristics and gear selectivity

No single conventional method for detecting species works across all habitats and organisms (Bonar et al., 2009, Neebling and Quist, 2011, Evans et al., 2017b). Studies that attempt to characterize species assemblages or identify relationships among habitat types and species occurrence, typically rely on the use of multiple types of sampling gear and approaches when conventional methods are deployed (Bonar et al., 2009). For example, seine netting cannot be used in rocky areas; electrofishing doesn't work in areas deeper than 2-3 meters; stable isotopes can provide information on movements at relatively broad scales, but require capture of target organisms and are dependent upon unique isotopic signatures existing in different locations where the organisms spend a portion of their life cycle (Cunjak et al., 2005). Seine netting and electrofishing are the most commonly used conventional methods, yet many fish are not vulnerable to capture by seine net and it has been shown that at least three passes with electrofishing gear are needed to obtain accurate biodiversity estimates (Kennard et al., 2006, Evans et al., 2017b). These approaches may require substantial investments of greater sampling effort and higher costs than eDNA methods (Evans et al., 2017b).

eDNA benefits and limitations

Data from habitat types that are notoriously difficult to sample, such as those present in large tropical rivers, may become accessible with eDNA analysis methods. Collection of water samples presents fewer logistical challenges than deployment of conventional gears in sites that are deep, rocky, full of debris, or have high flows. Data gathered by

collecting eDNA can provide critical information on the extent of species distributions and occupancy in habitat types that present logistical limitations on sampling with conventional gears (Bellemain et al., 2016). Researchers have effectively identified carp DNA in large canals linking the Mississippi River and Great Lakes basins, enabling early detection of this invasive species at low densities, which then informed rapid management action (Jerde et al., 2011). However, the literature on fish eDNA shows that more work is needed to validate and improve the method for specific habitats, biological variables, and environmental conditions.

Application of eDNA methods has shown that the volume of water collection needed to obtain sufficient molecular material to enable detection is highly variable and dependent on the size of the water body, abundance of target organisms, eDNA transport rates and DNA shedding rates (Sassoubre et al., 2016, Shaw et al., 2016, Shogren et al., 2017). Large tropical rivers have seen little attention in the eDNA literature (Doi et al., 2017), as they likely represent the most challenging and variable freshwater environments to survey with eDNA methods (Bellemain et al., 2016). Certain organic materials can inhibit some enzymes used in PCR amplification of eDNA, therefore, turbidity of sampling locations may influence detection results and testing for sample inhibition is recommended (Goldberg et al., 2016). Some research has shown that fish eDNA is more abundant in aquatic sediments than in surface water samples (Turner et al., 2015), while other research has shown better results from water collection when comparing water samples to sediment samples (Shaw et al., 2016).

2.4.3 Destructiveness of sampling

Reducing the potential for mortality caused by sampling is a high priority when studying rare species. Conventional methods can be destructive, both to target and non-target organisms, as well as the local environment (e.g. through habitat modification,

trampling etc.) (Olson et al., 2012). Fish sampling typically involves physical capture (e.g. using gill netting or electrofishing) and handling of animals with or without anesthesia, all of which can cause physical injury and even death to sampled fish. When the primary purpose of collecting data on rare taxa is to conserve them, it follows that every effort should be made to reduce the negative impacts of research on the survival of individuals. Rare species are often highly sensitive with small population sizes, and researchers have a responsibility to limit the destructiveness of their work.

eDNA benefits and limitations

Collection of eDNA is one of the most non-invasive biological sampling methods available, and has been applied to a broad range of terrestrial organisms that are extinct, highly endangered, sensitive, elusive, or otherwise difficult to survey (Beja-Pereira et al., 2009). These studies typically rely on feces, urine, hairs, feathers, skin, or eggshells that are detected in the environment after being shed by the target organism (Waits and Paetkau, 2005). For rare aquatic taxa, determining presence through collection of water samples rather than through physical capture of individuals highlights the potential for eDNA to provide a non-invasive solution, especially for sensitive organisms.

2.4.4 Taxonomic accuracy

Taxonomy of aquatic organisms is often unclear, and for many genera there are few experts with the skills needed to accurately identify specimens. For river systems with high species richness, cryptic species complexes are common. Conventional sampling methods typically rely on visual identification of species in the field. For some fishes the only way to distinguish them is to conduct extensive morphological analysis, such as counting fin rays and scale rows. This process is prone to error, time-consuming, and often requires expert lab analysis. For studies aiming to assess total biodiversity in

systems with high species richness, ensuring that all species present in a sample are properly identified can be very labor-intensive and produce inaccurate data.

eDNA benefits and limitations

Positive detections of a high number of species representing diverse taxa have been validated through laboratory and field experiments, where eDNA results were compared with conventional sampling (Hänfling et al., 2016, Shaw et al., 2016). Some studies show that the eDNA approach requires less effort than conventional sampling methods, and in some cases is more accurate in determining presence (Dejean et al., 2011, Goldberg et al., 2011, Mahon et al., 2013b). eDNA provides the advantage of distinguishing cryptic species at any life stage through genetic, rather than morphological, identification (Evans et al., 2017b).

While metabarcoding of eDNA generates a large amount of sequencing data, accurate taxonomic assignment of sequences requires reference sequences of the target species or their close relatives are available. If these are not currently available in online databases, they must be collected and sequenced from tissue samples of validated specimens. In many cases it is sufficient to use congeneric or confamilial taxa from databases to assign taxonomy through phylogenetic relatedness, assuming the short fragment of DNA between primer sites does not evolve too rapidly and become saturated. While it is possible to extract DNA from specimens in collections, many current metabarcoding studies are creating their own reference databases through collection of tissue samples from live individuals captured in the study area (Oliveira et al., 2016, Valentini et al., 2016, Cariani et al., 2017, Chang et al., 2017). For many species, reference sequences are not currently available for the appropriate gene regions and fragment sizes to accurately assign taxonomy at the species level to metabarcoding data.

2.4.5 Inferring absence

Field surveys can never determine absence at a specific site with absolute certainty. False negatives are always possible, no matter what type of sampling method is used, including eDNA. Conventional methods require statistical analysis to infer absence. False positives are unlikely with conventional methods, though they may arise from errors in taxonomic identification, while false negatives are quite common. A false negative is a dangerous error for highly threatened species, as it could result in the loss of an opportunity to protect habitat that may be essential for persistence.

eDNA benefits and limitations

There are indications that failure to detect a species using eDNA is a reliable gauge of species absence (Mahon et al., 2013b). The spatial and temporal resolution of data gathered with eDNA sampling are highly dependent on the environmental conditions which influence degradation rates of shed DNA. Research shows that eDNA tends to degrade rapidly in water, so a positive detection indicates the recent presence of a focal species while negative results are likely when the species is not present (Dejean et al., 2011, Thomsen et al., 2012b). However, false negative results can occur for many reasons, and improvements in processing samples and interpreting data are needed to better understand the cause of false negatives. Some research showed that eDNA may persist in the water column for approximately one to two weeks depending on environmental conditions (Dejean et al., 2011, Thomsen et al., 2012b), but degradation rates are known to be highly variable (Barnes et al., 2014).

For eDNA studies in lotic systems, the potential for DNA to be transported downstream can influence detection probabilities (Jerde et al., 2016, Shogren et al., 2017). Some work has shown that the transport distance where eDNA falls below detectable levels varies among species (Deiner and Altermatt, 2014). Careful design of field sampling, analysis of detection probabilities, and repeated sampling is required to

infer consistent absence with high confidence. Site occupancy models are an especially relevant tool, as they can be used to reduce the number of samples that are required for dependable estimates of absence. They use statistical inference to reliably estimate detection probabilities, even when detection is imperfect (Schmidt et al., 2013).

False positive detections are caused by factors such as contamination in the laboratory and bioinformatics procedures to assign taxonomy (Goldberg et al., 2016). These are unlikely to occur given standard controls and protocols that are observed in most eDNA studies. These protocols include physical separation of DNA extraction and PCR preparation from areas where PCR is performed, and use of negative controls throughout the entire molecular workflow. However, the lack of adequate reference sequences in online databases needs to be addressed to ensure appropriate taxonomy can be assigned to sequence reads from environmental samples.

2.5 Conclusions

A boom in published research on eDNA and the rapid advancement of DNA sequencing technology demonstrate the potential of the technique to enhance collection of data on rare species (Figure 2.2). However, this will require further work to overcome the current limitations described in this review. If and when the methodological challenges currently existing with application of eDNA methods can be overcome, the potential for conservation applications is extensive.

Species richness accounting, real-time species detection, revealing biogeography, more effective conservation interventions, and improved efficiency in legal actions regarding endangered species can all be advanced through the application and development of eDNA sequencing. Increasingly accurate estimates of species richness are accumulating through the application of eDNA methods and metabarcoding (Lodge et al., 2012, Thomsen et al., 2012b). At the same time, there is substantial room to

improve the methods and adapt them for different environments and species. Primarily, improvements are needed in the design of sampling strategies to ensure samples are representative of true species richness and that sequencing provides accurate and useful results.

In many developed countries, substantial legal actions can be triggered when an organism that is protected under endangered species legislation is present at a development site. eDNA methods can provide rapid and accurate presence/absence data for new project sites. These methods could be used to determine occupancy in proposed or designated critical habitats. eDNA has been used to detect spawning activity of an endangered fish in Australia (Bylemans et al., 2017). Routine monitoring using eDNA could be applied for adaptive management. For example, specific management actions, such as limiting water withdrawals, could be triggered only at the precise time when a protected migratory species moves to a specific area, not before and not after. This could reduce the economic burden of legislative requirements while also increasing the potential to effectively conserve endangered species. A study on the presence of an endangered fish (tidewater goby) compared seining results to eDNA results, and showed that eDNA analysis had higher detection probabilities (Schmelzle and Kinziger, 2016).

Systematic conservation planning and prioritization is a highly developed field, but faces limitations when data on species occurrence are not available, and when solely presence data are available, or when distributions are represented by just a few observations (Hermoso et al., 2014). The use of data collected through eDNA methods can help overcome these limitations. Such an approach can help to achieve the most conservation benefit out of areas prioritized by systematic planning models by validating the outputs using eDNA field surveys. More effective and efficient use of

conservation dollars can be achieved by increasing the ability to accurately identify priority sites with high confidence of presence of priority threatened species.

For metabarcoding studies, the approach and assumptions used in decisions about whether to retain or discard sequences have important implications for the accurate interpretation of results (Evans et al., 2017a). Rare species are less likely to be detected with a metabarcoding approach than with a species-specific approach, yet metabarcoding potentially offers information about entire community assemblages. If the goal of a study is to detect one or a few rare species, detection is more likely with the use of species-specific primers. Reference genomes are required to design such primers for target species. GenBank has the most comprehensive referenced list of available genetic sequences (Benson et al., 2012), and so most eDNA research relies on this database for taxonomic assignment of sequencing results. Currently, reference sequences for many aquatic taxa are limited, especially for rare species from the most biodiverse regions. Additional work on sequencing reference specimens that have been validated by experienced taxonomists is needed to alleviate this bottleneck.

Application of eDNA methods can give us more timely and accurate information on distributions of rare species, enabling truly adaptive management and improving the potential for conservation successes. With this method we can access dynamic distribution data, enabling the design of specific responses based on current or changing distributions. This information will become increasingly relevant as global climate change exerts its influence on distributions of organisms that may be desirable or invasive. Routine eDNA surveys in key habitat types can assist in monitoring shifts in richness or diversity metrics over time. The ability to understand dynamic distributions and to detect species at very low densities is an essential tool for effective conservation of rare organisms.

While conventional sampling methods will always be fundamental to aquatic species research, they are often not well-suited to the conditions and context associated with studying distributions of rare aquatic species. A complementary approach is needed for rapid assessments, to access locations that are inaccessible with conventional methods and to access data that enable more effective conservation of threatened taxa. If current limitations can be overcome, the tools built on eDNA and next-generation sequencing technology offer the potential for addressing the specific challenges encountered when sampling rare aquatic species. eDNA methods will not replace conventional methods, but they will likely provide a complementary approach that increases our ability to achieve targeted conservation outcomes. Understanding the potential of these tools and thorough consideration for appropriate ways to apply them will likely be highly advantageous. This potential will likely be realized in the near future and on broad scales, for scientists and managers tasked with understanding and sustaining populations of rare species

The following chapters of this thesis describe work to test eDNA methods and overcome some of the challenges described in this review, including: the role of sampling volume in species detection and ability to detect highly rare species in large tropical rivers; the influence of primer design on estimates of diversity and comparison of eDNA results with richness data obtained using conventional survey methods; the influence of environmental variability in river environments and the role of field sampling protocols in achieving representative estimates of richness; and finally, the potential for eDNA to be transported over long distances in a subtropical river basin and to be preserved in sediments to provide a chronology of taxa composition over long time periods

Chapter 3: TRAILS OF RIVER MONSTERS: DETECTING CRITICALLY ENDANGERED MEKONG GIANT CATFISH *PANGASIANODON GIGAS* USING ENVIRONMENTAL DNA

3.1 Statement of author contributions

Statement of contributions to published work

The following chapter of this thesis was published in a peer-reviewed journal and the text is included here as published. The work presented represents a collaboration between Griffith University researchers and professors, a private research company (SPYGEN, authors include AV and TD), and the World Wildlife Fund – Greater Mekong Program (WWF, authors include TG and FG). The doctoral candidate, Harmony Patricio (HP), is co-first author with a collaborator from SPYGEN, Eva Bellemain (EB). HP was responsible for project conceptualization, design and refinement of sampling strategy, and background research to determine sampling locations most likely to host the targeted species. She organized, managed and conducted all field work and sample collection in the Mekong River, including water sample collection for eDNA and interviews with local fishers to determine sampling points with highest probability for detection. Some samples were sent to SPYGEN for laboratory analysis as per a prior agreement between SPYGEN and WWF.

The bulk of original writing for the manuscript, approximately 80%, was done by HP. One author from an external research institute, CM, contributed to a final review of the manuscript. Professor Jon Olley (JO) and Professor David Lambert (DL) of Griffith University advised on conceptual design, sampling strategy, and field protocols. JO and DL contributed guidance to drafting of the manuscript and conducted extensive reviews. Funding was provided by JO, DL, and WWF. Some samples were tested and processed at the ancient DNA facility of Griffith University with the help of postdocs Leon Huynen (LH) and Richard O'Rorke (RO). Specific primers for Mekong giant catfish were

designed, tested, and validated by HP, LH, and RO. These data are not included in the manuscript, as the laboratory procedures were different from those applied by SPYGEN. Data analysis was conducted by SPYGEN and HP, with the help of JO, DL, and RO. This work is original and other than inclusion in this thesis, has only been published in the journal Global Ecology and Conservation.

Journal citation:

Bellemain, E., Patricio, H. C., Gray, T., Guegan, F., Valentini, A., Miaud, C., & Dejean, T. (2016). Trails of river monsters: Detecting critically endangered Mekong giant catfish Pangasianodon gigas using environmental DNA. *Global Ecology and Conservation, 7*, 148-156. doi: http://dx.doi.org/10.1016/j.gecco.2016.06.007

3.2 Summary

Pressures on freshwater biodiversity in Southeast Asia are accelerating, yet the status and conservation needs of many of the region's iconic fish species are poorly known. The Mekong is highly species diverse and supports four of the six largest freshwater fish globally, three of which, including Mekong giant catfish (*Pangasianodon gigas*), are Critically Endangered. Emerging environmental DNA (eDNA) techniques have potential for monitoring threatened freshwater biodiversity, yet have not been applied in complex and biodiverse tropical ecosystems such as the Mekong. We developed species-specific primers for amplifying Mekong giant catfish DNA. In situ validation demonstrated that the DNA amplification was successful for all samples taken in reservoirs with known presence of Mekong giant catfish independent of fish density. We collected water samples from six deep pools on the Mekong, identified through Local Ecological Knowledge, in Cambodia, Lao PDR, and Thailand. DNA was extracted and amplified from these samples using the designed primers and probes. Mekong giant catfish DNA was detected from one sample from the species' presumed spawning grounds on the Mekong mainstream, near the border between northern

Thailand and Lao PDR. eDNA sampling using species-specific primers has potential for surveying and monitoring poorly known species from complex tropical aquatic environments. However accounting for false absences is likely to be required for the method to function with precision when applied to extremely rare species that are highly dispersed within a large river system. We recommend that such approach be utilized more widely by freshwater conservation practitioners for specific applications. The method is best suited for baseline biodiversity assessments or to identify and prioritize locations for more rigorous sampling. Our methods are particularly relevant for systems or species with limited baseline data or with physical characteristics that logistically limit the application of conventional methods. Such attributes are typical of large tropical rivers such as the Mekong, Congo, or Amazon.

3.3 Introduction

Southeast Asia is facing a biodiversity crisis driven by unprecedented rates of growth in populations, economies, and per-capita consumption (Sodhi et al., 2010). This has led to a higher concentration of threatened terrestrial reptile, bird and mammal species than any other region globally (Sodhi et al., 2010). Patterns of freshwater biodiversity and endangerment in tropical Asia are less well known, but freshwater ecosystems often support a higher proportion of threatened species than marine or terrestrial equivalents (Loh et al., 2005, Revenga et al., 2005, Dudgeon et al., 2006, Strayer and Dudgeon, 2010). Given that trends in water use, energy production and consumption, and associated environmental degradation are projected to continue rising across Southeast Asia, understanding and monitoring the status of freshwater biodiversity is critical.

The accelerating pressures on freshwater ecosystems in tropical Asia are encapsulated by the Mekong; the largest river in Asia and ranked amongst the top three

rivers globally in terms of fish diversity (Campbell, 2009). Large stretches of the lower Mekong Basin (comprising Myanmar, Thailand, Lao PDR, Cambodia, and Vietnam) remain relatively pristine, but could be irreversibly altered by planned infrastructure developments. Many Mekong fish species are also currently impacted by heavy exploitation, particularly large-bodied and migratory species (Allan et al., 2005).

The Mekong supports a unique assemblage of freshwater megafauna (species > 90-kg and >180-cm long) including four of the six largest freshwater species globally (Hogan, 2011). Three of these species: Mekong giant catfish (*Pangasianodon gigas*), giant carp (*Catlocarpio siamensis*), and dog-eating catfish (*Pangasius sanitwongsei*), are assessed as Critically Endangered by the IUCN Red List, with declines driven primarily by heavy fishing pressure and habitat loss or degradation (Hogan, 2013b, Hogan, 2013a, Jenkins et al., 2009). The Mekong giant catfish is described as a contemporary example of overharvest (Allan et al., 2005).

Despite their notable role in regional culture, economies, and food security, knowledge of the current status and distribution of many charismatic flagships of Mekong freshwater biodiversity, including the Mekong giant catfish, is limited. This paucity of knowledge largely results from of a lack of effective survey and monitoring techniques. Because of population declines and associated fishing bans, the majority of recent (post-2005) records of Mekong giant catfish occurrence are from incidental by-catch by fishermen in Cambodia, Lao PDR, and Vietnam (e.g. Hogan et al. (2004)). There is thus an urgent need for effective survey methods for the species in order to both monitor distribution trends and to identify priority sites for conservation interventions.

Advances in molecular biology have provided means to monitor single species, or to assess community biodiversity, from aquatic or terrestrial environments using environmental DNA (eDNA), i.e. DNA left behind by the organism in the environment.

This approach has demonstrated potential to detect rare and cryptic species, with examples in the literature rapidly growing, and has been validated on different taxonomic groups (including vertebrates and invertebrates) and different ecosystems (Dejean et al., 2012, Taberlet et al., 2012a, Machler et al., 2014, Biggs et al., 2015, Thomsen and Willerslev, 2015). eDNA sampling of rivers in North America has been shown to enable cost-effective detection of fishes occurring at low densities, such as along invasion fronts(Jerde et al., 2011, Goldberg et al., 2015). These methods are opening new perspectives for improving environmental monitoring and informing management and policy (Kelly et al., 2014).

The majority of published applications of eDNA sampling have occurred in controlled conditions and/or in temperate environments with low species diversity (Thomsen and Willerslev, 2015). The applicability of eDNA sampling for effective conservation management in the biodiverse tropics is less clear. To test the efficacy of eDNA sampling for detecting threatened tropical species in a challenging environment, we developed, and independently verified, specific assays (primers and probes) for Mekong giant catfish. These assays were then applied to water samples collected in reservoirs with confirmed presence of Mekong giant catfish and from the Mekong River across the known range of the species.

3.4 Materials and methods

3.4.1 Design, in silico and in vitro validation of an assay for Mekong giant catfish

We designed species-specific primers and a probe for amplification of Mekong giant catfish DNA. All available cytochrome b sequences for the Mekong giant catfish and other species from the Pangasiidae family 16 out of the 28 species from the genera *Heicophagus*, *Pangasianodon*, *Pangasius* and *Pseudolais* (FishBase, 2016), were retrieved from GenBank® and then aligned using Geneious® software (version R6,

http://www.geneious.com) (Kearse et al., 2012). The PrimerQuest® program (IDT, Coralville, USA. Retrieved 12 December, 2012. http://www.idtdna.com/Scitools) was used to design primers and probes. The designed primers were first validated in silico by assessing their specificity using the ecoPCR program (Bellemain et al., 2010, Ficetola et al., 2010) on the EMBL-Bank release 114 (released in December 2012). An in vitro validation was then undertaken using real time PCR upon 10 Mekong giant catfish DNA samples, extracted from captive fish in Thailand. The limit of detection (LOD, i.e. the minimum amount of target DNA sequence that can be detected in the sample) was calculated by running a dilution series of a known amount of DNA, ranging from 10^{-1} ng/µL to 10^{-10} ng/µL. The real time amplification (qPCR) was performed in 12 replicates in a final volume of 25 µL, using 3 µL of template DNA, 12.5 µL of TaqMan Environmental Master Mix 2.0 (Life Technologies®), 6.5 µL of ddH2O, 1 µL of Forward primer (Pgigas_cytb_F:CTAACCTGGATTGGTGGCAT, 10 µM), 1 µL of Reverse primer (Pgigas_cytb_R: AAGAAGAGGAAGTACAAGATGGAG, 10 µM) and 1µL of probe (Pgigas_cytb_Pr: CCAATAATGATGAATGGATGTTCGACTGGC, 2.5µM) under thermal cycling 50 °C for 5 min and 95 °C for 10 min, followed by 50 cycles of 95 °C for 30 s. and 58.8 °C for 1 min. qPCR negative controls (with 12 replicates as well) were performed in parallel to detect potential contamination. Samples were run on a BIO-RAD® CFX96 Touch real time PCR detection system in a room dedicated to amplified DNA analysis with negative air pressure and physically separated from the DNA extraction room.

3.4.2 In situ validation of Mekong giant catfish assay

In situ validation of Mekong giant catfish specific assay was undertaken using eight water filtration samples collected from ponds of the Srichiangmai Fisheries Research Centre in Nong Khai Province, Thailand, which contained captive Mekong giant catfish at different densities, from ''low'' to ''high'' (Table 3.1).

Sampling was performed from the shore using, for each sample, a filtration capsule (Envirochek HV®1 μm, Pall Corporation, Ann Arbor, MI, USA), sterile tubing and joints and a portable peristaltic pump (Vampire Sampler, Bürkle GmbH®, Germany) with a flow of 1.67 L/min. For each sampling, 10 liters were filtered near the surface (maximum depth 40 cm). After filtration, 125 mL of ethanol was added to each capsule, and the capsules were kept refrigerated at 4 °C for a few days until sent to the laboratory. The transport to the laboratory took 4 days. DNA extraction was conducted, following a modified protocol from Valentini et al. (2016), in a room dedicated to processing water samples equipped with positive air pressure, UV treatment and frequent air renewal. Laboratory personnel wore full protective clothing (disposable coveralls, hood, mask, laboratory-specific shoes, overshoes and two pairs of gloves) that was put on in an airlock foyer before entering the processing room. Filtration capsules were emptied of the ethanol and filled with 125 mL conservation buffer (Tris–HCl 0.1 M, EDTA 0.1 M, NaCl 0.01 M and N-lauroyl sarcosine 1% with pH 7.5–8). They were then left at 56 °C for 2 h, agitated manually for 5 min and emptied into three 50 mL tubes. In total, approximately 120 mL were retrieved in three tubes that were centrifuged for 15 min at 15,000 g. The supernatant was removed with a sterile pipette, leaving 15 mL of liquid at the bottom of the tube. Subsequently, 33 mL of ethanol and 1.5 mL of 3M sodium acetate were added to each 50 mL tube. After a quick manual shaking, samples were stored for 24h at −20 °C. The three subsamples per site were then centrifuged for 30 min at 15,000 g and 6 °C and the supernatant was discarded. After this step, 360 μL of ATL Buffer of the DNeasy Blood & Tissue Extraction Kit (Qiagen) were added in the first tube, the tube was vortexed and the supernatant was transferred to the second tube. This operation was repeated for the third tube. The supernatant in the 3rd tube was transferred in a 2 mL tube and the DNA extraction was

performed following the manufacturer's instructions. Two mock extractions were performed in parallel to monitor possible contaminations. Real time DNA amplifications (including mock samples) were carried out in 12 replicates using the specific primers designed in this study for Mekong giant catfish, following the protocol described above. qPCR negative controls (12 replicates) were performed in parallel to detect potential contamination. A dilution series of Mekong giant catfish DNA, ranging from 10-1 ng/µL to 10-4 ng/µL, was used as a qPCR standard. Those standards were added in the qPCR plate (with the tubes containing the eDNA samples sealed) in a separate room from the one where the eDNA extractions were performed.

3.4.3 Selection of sampling sites in the Mekong

Mekong giant catfish is endemic to the lower and middle Mekong basin and was historically distributed throughout the Mekong from the coast of Vietnam to northern Lao PDR (Hogan, 2013b). Whilst understanding of the species' migration patterns is incomplete, it is believed to migrate along the Mekong and Tonle Sap River, between spawning sites in northern Thailand and Lao PDR, to nursery grounds in the Tonle Sap Lake, Cambodia. Water sampling was undertaken at six locations in the Mekong between the confluence of the Tonle Sap and Mekong, close to Phnom Penh in Cambodia, to northern Thailand (Figure 3.1).

Broad sampling locations were selected based on historic catch records indicating presence of Mekong giant catfish, Local Ecological Knowledge (LEK) on the species' distribution, as well as logistics and accessibility constraints. LEK data was sourced during previous projects for Mekong giant catfish conservation, as well as informal interviews with experienced fishermen. Considering that Mekong giant catfish are primarily a benthopelagic species relying on deep pools as refugia, especially in the dry season when river levels are low (Poulsen et al., 2002), we targeted sampling at

deep pools using hydrographic maps produced by the Mekong River Commission combined with informal interviews with senior fishermen who identified specific pools where the species had been captured in the past (Halls et al., 2013a, MRC, 1996). Sampling was performed in April and May 2014, at the end of the dry season, with relatively low river flow compared to the rest of the year.

SAMPLE NAME	DENSITY OF CATFISH	NUMBER POSITIVE REPLICATES
Res1 n°1	Medium	12/12
Res1 n°2	Medium	12/12
Res1 n°3	Medium	12/12
Res2 n°4	Low	12/12
Res2 n°5	Low	12/12
Res2 n°6	Low	12/12
Res3 n°7	High	12/12
Res3 n°8	High	12/12
Negative	Species absent	0/12

Table 3.1: Number of samples with positive detection of Mekong giant catfish from aquaculture ponds in Thailand during In situ validation of species-specific assay. Densities of catfish in each reservoir defined as follows: High: ±12 MGC – wild individuals captured in the Mekong at least 20 years ago (‘‘the parents’’) (40–50 kg weight), in a pond of 3200 m^2 with a 2–3 m depth; Medium: ±40 large MGC (40–60 kg) – some up to 20 years old + about 100 smaller MGC of less than 10 kg (F1, second generation bred from wild parents), in a pond of 24, 000 m^2 with a 4 m depth; Low: ±10 MGC – 25 cm length (1–2 kg) in a pond of 6400 m^2 with a 2–3 m depth.

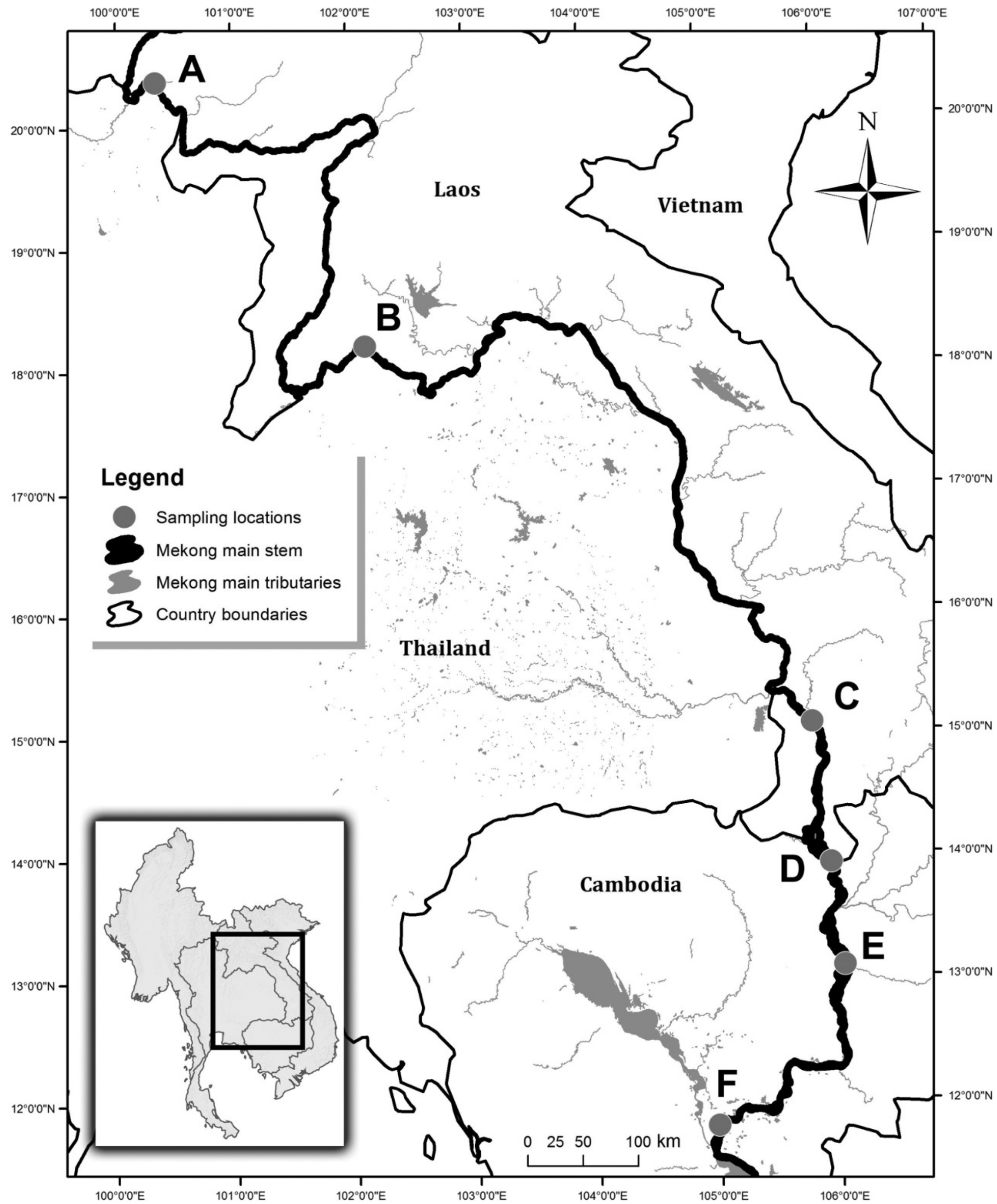

Figure 3.1. Sampling sites for Mekong giant catfish along the Mekong River in Cambodia, Lao PDR, and Thailand. More details of each sampling site are given in Table 3.2

Site	Sample number	Site	Location	Date (2014)	GPS N	GPS E	Depth (m)	Type
A	A1	Upstream of Ban Pla Tang	Deep pool	13/05	20′21.892	100′21.434	0	Surface
	A2	Upstream of Ban Pla Tang		13/05	20′21.892	100′21.434	26.65–0.91	Water column
	A3	Upstream of Ban Pla Tang		13/05	20′21.892	100′21.434	29.57	Bottom
	A4	Upstream of Ban Pla Tang	Up-stream	13/05	20′23.111	100′20.652	0	Surface
	A5	Upstream of Ban Pla Tang		13/05	20′23.111	100′20.652	24.08–0.91	Water column
	A6	Upstream of Ban Pla Tang		13/05	20′23.111	100′20.652	24.99	Bottom
B	B1	Nong Pla Beuk	Up-stream	22/04	18′12.716	102′07.625	6.10	Bottom
	B2	Nong Pla Beuk		22/04	18′12.716	102′07.625	5.1–1	Water column
	B3	Nong Pla Beuk		22/04	18′12.716	102′07.625	0	Surface
	B4	Nong Pla Beuk	Deep pool	22/04	18′12.501	102′07.562	28.65	Bottom
	B5	Nong Pla Beuk		22/04	18′12.501	102′07.562	27.65–1	Water column
	B6	Nong Pla Beuk		22/04	18′12.501	102′07.562	0	Surface
C	C1	Pakse	Up-stream	05/05	15′04.485	105′49.851	0	Surface
	C2	Pakse		05/05	15′04.485	105′49.851	17.29–1	Water column
	C3	Pakse		05/05	15′04.485	105′49.851	18.29	Bottom
	C4	Pakse	Deep pool	05/05	15′05.586	105′49.692	0	Surface
	C5	Pakse		05/05	15′05.586	105′49.692	26.43–1	Water column
	C6	Pakse		05/05	15′05.586	105′49.692	27.43	Bottom
D	D1	Dolphin pool	Up-stream	02/05	13′55.939	105′57.524	0	Surface
	D2	Dolphin pool		02/05	13′55.939	105′57.524	21.86–1	Water column
	D3	Dolphin pool		02/05	13′55.939	105′57.524	22.86	Bottom
	D4	Dolphin pool	Deep pool	02/05	13′56.033	105′56.936	0	Surface
	D5	Dolphin pool		02/05	13′56.033	105′56.936	35.58–1	Water column
	D6	Dolphin pool		02/05	13′56.033	105′56.936	36.58	Bottom
E	E1	Koh Khnea	Up-stream	01/05	13′05.675	106′02.912	0	Surface
	E2	Koh Khnea		01/05	13′05.675	106′02.912	12.72–1	Water column
	E3	Koh Khnea		01/05	13′05.675	106′02.912	13.72	Bottom
	E4	Koh Khnea	Deep pool	01/05	13′07.746	106′03.506	0	Surface
	E5	Koh Khnea		01/05	13′07.746	106′03.506	14.24–1	Water column
	E6	Koh Khnea		01/05	13′07.746	106′03.506	15.24	Bottom
F	F1	Phat Sanday	Up-stream	29/04	11′48.207	104′58.773	0	Surface
	F2	Phat Sanday		29/04	11′48.207	104′58.773	17.9–1	Water column
	F3	Phat Sanday		29/04	11′48.207	104′58.773	18.9	Bottom
	F4	Phat Sanday	Deep pool	29/04	11′50.679	104′59.703	0	Surface
	F5	Phat Sanday		29/04	11′50.679	104′59.703	19.73–1	Water column
	F6	Phat Sanday		29/04	11′50.679	104′59.703	20.73	Bottom

Table 3.2: Details of sampling sites surveyed during field surveys collecting water samples for Mekong giant catfish eDNA analysis on the Mekong River

3.4.4 eDNA sampling in the Mekong

At each site (n = 6, Figure 3.1), samples were collected from a boat in vertical series at two points within the pool: first at the downstream lip of the deep pool, and second, at the deepest point in the pool. At each sampling point, an immersion cylinder (Diameter 75 mm, capacity 1L; Bürkle GmbH® , Germany) was used to collect the water at different depths: (1) near-bottom, (2) water column, and (3) surface (Table 3.2).

The maximum depth at the sampling point was measured with a depth sounder. The near-bottom samples were collected one meter above the substrate. For the mixed

water column samples we subtracted two meters from the total depth (so that the sampling covered the column starting 1 m above the substrate up to 1 m below the surface) then divided this depth by 10. One liter of water was collected at each of the 10 depth points along the water column. Before each sampling round, immersion cylinders and buckets were sterilized with 10% chlorine bleach and a new sterile tube was used for each water filtration. At each depth, 10 liters were collected and emptied into a sterile bucket. Immediately after water collection, filtration was performed from the buckets, using the same filtration equipment as described for aquaculture ponds in Thailand. Capsules were then filled with conservation buffer (Tris–HCl 0.1 M, EDTA 0.1 M, NaCl 0.01 M and N-lauroyl sarcosine 1% with pH 7.5–8) and kept on ice in a cooler and then refrigerated at 4 °C until sent to the laboratory. Personnel wore gloves at all times during the sampling process. DNA extraction and amplification were performed following the protocol described in ''In situ validation of Mekong giant catfish assay'' section. In total, 36 water samples were collected, filtered and analyzed (30 liters filtered for each sampling point, and 60 liters filtered for each deep pool).

3.5 Results

3.5.1 In silico, in vitro, and in situ validation of the assay for Mekong giant catfish

The amplified fragment length was 83 bases, which is appropriate for amplification of environmental DNA that is usually degraded (Deagle et al., 2006). The results of the in silico test showed that, using this assay, no other species present in the Mekong was amplified. However; four species of Micropterus (*M. cataractae*, *M. notius*, *M. punctulatus* and *M. treculii*) and one species *of Rineloricaria* (*R. uracantha*), all occurring in the continental United States, could potentially be amplified with 2 or 3 mismatches on each primer. When using the probe in combination with the primers,

only Mekong giant catfish is amplified. The in vitro tests showed that limit of detection of primers and probe was 5.10−8 ng of DNA/µL. The in situ tests (from aquaculture ponds containing Mekong giant catfish in Thailand) confirmed the reliability of the assay. For all samples collected in those ponds, the DNA amplification was successful, with all PCR replicates positive independent of fish density in the reservoirs (Table 1). The field, extraction and PCR negative controls did not show any amplification, validating the primer pair and the probe, confirming the purity of the consumables used and the absence of cross contaminations.

3.5.2 Field survey results

Among the 36 samples analyzed over the 6 sites, one sample (site A in Figure 3.1; sample A5 in Table 3.2), corresponding to a water-column sample, showed a positive amplification of Mekong giant catfish, in one replicate out of 12. Extraction and PCR negative controls did not show any amplification

3.6 Discussion

Detection and monitoring of elusive and rare species has been a long-standing challenge for conservationists. This challenge is particularly acute in environments such as large tropical rivers, which typically host highly threatened species in low densities and are not amenable to conventional survey methods. Environmental DNA (eDNA) sampling offers an innovative approach to overcome the challenges associated with detecting rare species in freshwater ecosystems (Goldberg et al., 2011). However, to our knowledge, there are no published studies using eDNA sampling in biodiverse tropical freshwater environments to target poorly understood threatened species. While eDNA methods are rapidly gaining popularity, with many publications bolstering the conception of the approach as a ‘‘silver bullet’’ for rare species surveys (Thomsen and

Willerslev, 2015, Valentini et al., 2016), our study serves to demonstrate both the potential and the limitations of these methods.

Here we demonstrate that by using eDNA barcoding with a specifically designed genetic assay, it is possible to detect the critically endangered Mekong giant catfish from reservoirs where the species is known to be present. Our detection in the natural habitat of this species was limited to a single positive detection, from the only known spawning grounds on the Mekong mainstream at the border between northern Thailand and Lao PDR. When we consider the scale of the river relative to the extremely low density of the species, it still represents a promising outcome. Our limited results from the natural habitat of the giant catfish do not reflect a problem with the molecular methods underlying eDNA research, rather they highlight the rarity of Mekong giant catfish eDNA in the sampled water. In addition, it shows the importance of obtaining prior knowledge of detection probabilities, rigorous statistical design of sampling strategies, and accounting for 'false absences' through analytical approaches such as occupancy modeling (MacKenzie et al., 2002).

Despite supporting high species richness, including more species of very large-bodied fish and more total fish biomass than any other river system (Baran et al., 2012), and providing critical ecosystem services, there is little systematic monitoring of freshwater biodiversity in the Mekong. Thus, the status of many iconic threatened species is poorly known. Currently available information is primarily based on interviews with fishermen, self-reporting of catches by fishers, or sampling of fish captured through commercial or subsistence harvests (Bao, 2001, Halls et al., 2013b, Fily and d'Aubenton, 1965). These conventional methods have limited potential to detect rare species, and the information they provide can be inaccurate or incomplete, and time-consuming or costly to collect.

Conventional aquatic species survey methods that are commonly used in North America and Europe, such as electrofishing or seine netting, are rarely applied at an adequate scale for research in the Mekong because of the large geographic scale and transboundary nature of the catchment, large number of long-distance migratory fish species, financial constraints, and logistically challenging physical characteristics of the river. Deep pools in the Mekong can reach depths up to 70-m, which makes the application of conventional methods impossible in many sites. One alternative option for monitoring species in deep pools is sonar sampling (Viravong et al., 2006). However, this method typically only provides information on biomass and does not enable species identification. Given these knowledge and methodological constraints eDNA sampling may be the most effective way to detect presence of threatened species in many sites, particularly if combined with preliminary site selection based on Local Ecological Knowledge (LEK).

3.6.1 Factors influencing detection probability of the giant catfish and recommendations for future eDNA sampling

Despite sampling throughout the known historic range of Mekong giant catfish across three countries, we obtained the species' DNA from only one of 12 replicates in the water column sample at the Ban Plan Tung sampling site (site A, Figure 3.1). Intrinsically, the detection probability of rare species will be low (Yoccoz et al., 2001, Kéry and Schmidt, 2008). The probability of collecting eDNA in any given river water sample is influenced by a cascade of events:

First, the species is or was recently present at the sampling location or some distance upstream. In this study, we used LEK to target sampling sites for eDNA collection in order to increase the chances of detecting our focal species. The literature contains highly variable results regarding the downstream transport distance of eDNA in rivers, and the distance downstream at which an organism's eDNA is detectable

seems to differ amongst species (Deiner and Altermatt, 2014, Pilliod et al., 2014, Civade et al., 2016). There are currently no published studies available which examine this process in a river as large as the Mekong. Having demonstrated proof-of-concept of detecting rare species using species-specific primers, we recommend that future studies design sampling using a site-occupancy framework (*sensu* Schmidt et al. (2013)), with multiple samples collected within short periods of time from sampling sites, to account, and model, for non-detection. Such an occupancy sampling approach could form a framework for long-term monitoring of Mekong giant catfish and may be particularly important given detection probability is certainly low.

Second, when the species is present at the sampling site a sufficient quantity of its DNA has to be released at the time of sampling. Density of the released DNA depends on the excretion rates of the target species, which varies among species (e.g. Klymus et al. (2015)), life stages (for example, the release rate was 3–4 times higher in adult than in juvenile fish (Maruyama et al., 2014)), and metabolic rates. No data is available on the DNA excretion rates of the Mekong giant fish, but the experiment performed in the ponds of the Srichiangmai Fisheries Research Centre (above) shows that the giant catfish releases sufficient DNA for eDNA detection. However fish density in such aquaculture ponds is usually much higher than fish density in natural environments. Future work to quantify DNA shedding rates for different genera or functional groups is needed to improve the accuracy, sensitivity and precision of eDNA surveys in the field.

Finally, released eDNA has to be collected in water samples. The probability of this occurring depends on the concentration of released eDNA, its persistence/degradation rate, its vertical, lateral, and longitudinal spatial distribution, and on the sampling strategy (Pilliod et al., 2014). The amount of eDNA at a site can be positively correlated with the target species density (Takahara et al., 2012, Mahon et al.,

2013a, Eichmiller et al., 2014, Moyer et al., 2014), yet the ratio of fish density to eDNA concentration is not consistent among different species or taxa (Kelly et al., 2014). Maruyama et al. (2014) showed that the half-life of fish eDNA in tanks after removal of live fish was 6.3 h, and Dejean et al. (2011) showed that detection probability was null two weeks after removal of animals stocked in ponds. The primary factors which would cause a marked difference in these rates in the Mekong when compared to temperate systems are temperature, UV light exposure, and levels of microbial digestion of shed DNA. Given the well-established influence of environmental factors on DNA degradation, we would expect persistence rates to be lower in a tropical setting. Moyer et al. (2014) used experimental ponds to demonstrate that for every 1.02 °C increase in temperature, the per-L-sample probability of eDNA detection decreased by 1.67 times. It is unclear whether this relationship would hold in a dynamic river system, where nearly all environmental variables and their influence on each other are markedly different than in ponds. The obvious and likely most influential difference is the high variability in water volume at any given site as a function of river velocity, discharge, and channel morphology.

Although the volume of water collected in eDNA studies is usually low (less than 2 L) it has been demonstrated that this volume exerts a significant and positive influence on detection probability. Schultz and Lance (2015) modelled the sensitivity of eDNA methods based on data from carp (Hypophthalmichthys spp.) and showed that increasing sampling volume from 2 to 3 L contributes the most to increasing sensitivity for field surveys. The logistical challenges presented by the need to filter very large volumes of water to detect rare species in large rivers warrant additional research, and there is plenty of room to adapt/improve field methods to function in this context. In addition, the velocity and discharge of a river or sampling site will also influence the

movement and concentration of eDNA once it is shed, and eDNA can persist in detectable levels over relatively short distances in streams, e.g. few km for fish (Civade et al., 2016).

Mekong giant catfish is thought to be benthopelagic (Poulsen et al., 2002) so it is interesting to note that the positive detection occurred in the water column sample and not in the water collected at the bottom of the deep pool. Perhaps the water column offers the best detection probability because it includes water that has been mixed from different depths. In a system with complex bathymetry like the Mekong it is possible for unique currents and upwelling to exist at the scale of a single sampling site. The sampling strategy for the Mekong giant catfish was based on the filtration of 10 L of flowing water at each sampled depth, with a total of 30-L per sampling point (n = 2 per site) within each pool, and a total of 60-L per pool. To improve upon our limited results new studies have to be designed to evaluate the efficiency of the sampling strategy according to the expected concentration of Mekong giant catfish eDNA in the river under different animal density and river discharge scenarios.

Finally when eDNA is successfully collected within the water sample, its detection will depend on several factors (e.g. Success of DNA extraction, PCR amplification sensitivity), (Wilcox et al., 2013, Ficetola et al., 2015). Sample collection, extraction, dilution, PCR, and sequencing methods can all influence results (Deiner et al., 2015, Schultz and Lance, 2015, Eichmiller et al., 2016b). In summary, the obtained result – the detection of the Mekong giant catfish in one of six sampled locations – has to be considered as the demonstration of the potential use of the eDNA technique to search for an extremely rare fish species in a large tropical river, but not as a full Mekong giant catfish survey. In order to increase chances to retrieve DNA from the target species in the future, improvements at different steps of the protocol (e.g. eDNA sampling effort

and strategy) should be considered. For instance, an integrative filtration strategy, as proposed by Valentini et al. (2016), could be used.

3.6.2 Perspectives for the conservation of the giant catfish and other Mekong fauna

Ban Plan Tung, the site where we detected Mekong giant catfish eDNA, is close to the only known spawning site of the species. Annual ceremonial fishing around the spawning season occurred at this site until the early 2000s (Hogan, 2013b). Population genetics studies have shown that giant catfish migrate at least 1000kmfrom the Tonle Sap in Cambodia upstream to the only known spawning site near Ban Plan Tung, Thailand (Ngamsiri et al., 2007). Unfortunately, this site is upstream of the first mainstream dam being constructed on the lower Mekong near Xayaburi, Lao PDR. Several mainstream dams have already been built in China, but these are upstream of the known distribution of the Mekong giant catfish. It is unclear whether mitigation measures that are associated with the construction of the Xayaburi dam will allow passage of Mekong giant catfish, and other large migratory fishes. Once the dam has been completed future eDNA surveys would allow us to see if the known spawning site upstream of the infrastructure is still occupied. eDNA sampling could be also be used to monitor the 234 fish conservation zones (FCZs) or ‘‘no-take’’ fish sanctuaries that have been established throughout Lao PDR. Targeted eDNA surveying could assess whether Mekong giant catfish utilize these conservation zones and we could maximize the efficient distribution of FCZs throughout the country for more effective species conservation (Baird and Flaherty, 2005).

Given the limited availability of baseline biodiversity data (i.e. little a priori knowledge of species composition) for many locations in the Mekong catchment, including the FCZs, eDNA metabarcoding (Taberlet et al., 2012a) would be a particularly useful tool for this region. This approach, based on high throughput sequencing, has

recently been developed and validated for the monitoring of aquatic biodiversity of fishes and amphibians (Valentini et al., 2016). It allows obtaining a list of taxa present in the sampled environment. Simmons et al. (2015) applied the term ''passive molecular surveillance'' to this metabarcoding approach and underlined its capability of detecting unexpected invasives. Unfortunately, fish species from tropical river systems as the Mekong are not well represented in public reference databases such as GenBank® and further sequencing efforts need to be carried out to alleviate this pitfall. Alternatively, in the absence of species identification, it is possible to work with Molecular Taxonomic Units (MOTUs) and to monitor those MOTUs in space and time. The discovery of undescribed fish taxon will probably motivate research for taxonomists and ecologists. We propose that such multi-species approaches, combined with species-specific assays as used in this study, have the additional capacity to detect rare and secretive species. We emphasize that accurate interpretation of results and avoidance of false negatives requires thorough analysis of detection probabilities and sensitivities under specific sampling scenarios and molecular workflows. We expect this approach will be increasingly utilized by freshwater conservationists to improve knowledge of aquatic biodiversity, particularly in poorly understood systems such as the Mekong.

Chapter 4: BIAS IN TAXON RICHNESS AND DIVERSITY ESTIMATES WHEN METABARCODING ENVIRONMENTAL DNA FROM RIVER WATER WITH THREE CLADE-SPECIFIC BONY FISH PRIMER SETS

4.1 Summary

Molecular survey methods are increasingly being used to provide biodiversity data for aquatic environments. Biases from PCR and sequencing processes are known to influence metabarcoding results, but it is assumed that this applies equally across samples. Here we examine this assumption and find that taxon composition and beta diversity estimates vary across environmental DNA (eDNA) samples when amplified with three different universal bony fish (Osteichthyes) primer sets. The number of molecular operational taxonomic units (OTUs) resulting from the use of each primer set was 71, 24, and 35 for the three markers, respectively. Water samples were collected from the Mary River, Queensland, Australia in conjunction with fish collection by seine netting. Taxon composition and beta diversity estimates from metabarcoding of eDNA were compared to composition data from extensive prior sampling in the river using conventional capture methods and from site-specific predictions of occurrence using species distribution modeling. Two genera observed with seine netting were not detected by eDNA analysis, while eight genera that were not observed with seine netting were detected through eDNA metabarcoding. Composition of taxa identified by each clade-specific fish primer was not uniform among primer sets. Results indicate that if the goal of an eDNA metabarcoding study is to accurately assess taxon composition and calculate biodiversity indices, multiple primers or benchmark/control data sets should be used to improve accuracy. This study suggests that while the potential utility of eDNA metabarcoding is promising, further work is needed to refine the method to a point where it is suitable for broad application in conservation and surveying. Some of the most pressing issues to be addressed include: improving quantification of detection

probabilities, advancing understanding of the influence of environmental variability at field sites on probability of collecting target eDNA, and increased effort to produce reference sequences for highly biodiverse regions of the globe.

4.2 Introduction

High-throughput sequencing of environmental DNA (eDNA) with generic primers ('metabarcoding') is increasingly being used to estimate aquatic species composition, richness and diversity (Thomsen et al., 2012a, Miya et al., 2015, Valentini et al., 2016, Olds et al., 2016). Several recent reviews have suggested that these eDNA methods are more sensitive and feasible than conventional sampling methods (Rees et al., 2014, Thomsen and Willerslev, 2015). However, the process of amplifying and sequencing DNA is subject to errors which may bias richness and diversity estimates (Dickie, 2010). Only a few published papers have quantified this bias or tested whether it is consistent across samples or primer sets (Tang et al., 2012, Zhan et al., 2014, Ficetola et al., 2015). Aquatic eDNA samples present unique challenges because a large proportion of the sample will be derived from non-target DNA (i.e. plant or bacteria). Despite attempts to amplify target to isolate a particular locus, a large proportion of sequencing data from eDNA samples often come from non-targeted loci and are discarded after quality filtering. For example, Miya et al. (2015) found that 42.3% of sequences came from unknown sources when metabarcoding environmental samples of seawater. eDNA that is relatively rare is less likely to be detected because of limited sequencing depth, and without control experiments it is not possible to discern whether particular eDNA is rare because the organism is rare, because it sheds less DNA, or because of PCR and sequencing biases (Kelly et al., 2014). Common species may swamp out the detection of rare amplicons in sequencing, which may erroneously indicate lower richness levels (Adams et al., 2013).

This chapter attempts to further understanding of this challenge by comparing multiple universal bony fish primers simultaneously to determine the level of bias in results from metabarcoding of fish eDNA from water samples. To assess the reliability of metabarcoding results, they were compared to several alternative sources of species occurrence data, including: seine netting at eDNA sampling sites, a reference list of species known to occur in the river, and species distribution models.

eDNA metabarcoding is the batch-processing of complex environmental samples containing DNA that has been shed by a wide variety of organisms using universal markers (Taberlet et al., 2012a), and has recently become relatively affordable because of improvements in DNA sequencing technology (Stein et al., 2014). Universal molecular markers are designed to maximize taxonomic coverage of polymerase chain reaction (PCR) amplification within a specific group (e.g. bony fish) and provide precise taxonomic resolution, while minimizing the amplification of taxa outside of that group (e.g. reptiles) (Folmer et al., 1994). However, eDNA studies typically rely on short, degraded fragments of DNA, which means they are especially prone to errors caused by differences in amplification and sequencing efficiency (Sefc et al., 2007). Primers may preferentially amplify DNA of some species over others, and eDNA belonging to species present in low densities at sampling sites may be swamped in sequencing by highly abundant species. Drop-out of targeted sequences and variation in amplification and sequencing efficiency can result in richness estimates that may be skewed to over/underestimate occurrence of particular taxa over others.

The use of multiple markers with different target loci may improve accuracy of richness estimates, by reducing drop-out rates or buffering the influence of preferential amplification by a single primer set (Zhan et al., 2014, Olds et al., 2016). Accurate information on species composition and its spatiotemporal variation is essential for

developing ecological understanding and effective conservation and management strategies for threatened ecosystems and biodiversity.

Beta diversity is a means of reducing the complexity regarding the ways that samples differ into a single numerical value. When applied to sampling units in a specified area, beta diversity measures the variance in species composition among units (Anderson et al., 2006). For each sample, a dissimilarity matrix compares the difference in the number of sequencing reads for each assigned taxa (or the presence/absence of a specific taxa) to determine relative diversity values. It is expected that PCR and sequencing biases for metabarcoding data occur equally across all samples for a single marker. Following this expectation, between-sample biodiversity metrics (i.e. beta diversity) should be comparable. Because the value being compared between sites is the measure of dissimilarity between sites, it is assumed that equally-applied biases will not skew estimates of relative dissimilarity. The assumption that biases are equally applied across samples can be tested through the use of multiple primer sets, and comparison of the relative dissimilarities in diversity estimates produced by each primer set.

This study tests two primary hypotheses: (1) the use of multiple universal primers will overcome drop-out of targets from sequencing and amplification biases to provide more accurate taxon composition estimates than would a single universal primer, and (2) PCR and sequencing biases occur equally across samples, therefore β diversity will be consistent among primer sets. A universal primer set for bony fishes (*Osteichthyes*) was designed and validated for the 12S mitochondrial gene (“Ostei12S”). Two additional universal bony fish primers were taken from published literature (Miya et al., 2015, Valentini et al., 2016). Fish genus richness estimates obtained by eDNA metabarcoding of river water with three universal primer sets were compared to each

other. The accuracy of metabarcoding results was assessed through comparisons with three alternative sources of taxon composition data: a reference list of species known to occur in the sampling area, conventional sampling (seine-netting) conducted simultaneously with eDNA sampling, and predictions of taxon composition from species distribution models.

4.3 Methods

4.3.1 Field Methods

All sampling was conducted in March-April 2015 at four locations along the Mary River in southeast Queensland, Australia (Figure 4.1). Each eDNA sample was collected from the riverbank, with average water depth of 0.8m. eDNA sampling was conducted prior to any personnel or conventional sampling equipment entering the water. At each sampling site two 1-L autoclaved plastic bottles (Nalgene®) were submerged just below the water surface until full, with 1L collected first at the downstream end of the site and another 1L collected approximately 20m upstream, at the upstream end of the site. A "site" was defined as a single hydraulic unit. Bottles were stored on ice (maximum 7 hours) until filtration using a vacuum pump and flask. Each 1-liter sample was filtered through a separate glass fiber filter (2.4 μm pore diameter, Whatman®), filters were placed in sterile vials with 3000 μL of SET buffer with 0.5% SDS, and kept on ice until arrival at the lab, where they were stored at -80°C until extraction. All equipment was sterilized in an autoclave or with a 1.5% Sodium Hypochlorite solution and gloves were worn at all times. Following eDNA sampling, conventional fish sampling was conducted for another study on the same days and in the same locations as eDNA sampling using a seine net (30m long x 2m deep, 11mm mesh size). Fish collected by seine netting were identified to species following Allen et al. (2003).

4.3.2 Laboratory Methods

Primer design and validation was conducted with Geneious software version 8.1.1 (Biomatters, Ltd), Primer-BLAST (Ye et al., 2012), reference mitochondrial genomes for bony fishes from NCBI's GenBank (Benson et al., 2012, NCBI), and tissue extracts from several species of freshwater and marine fish. Standard PCR was performed and the product was visualized on an agarose gel to check for an amplified product of the expected length. The designed primer sequence targeted a region of the same gene (12S) as two additional primer sets that were selected from the published literature (Table 4.1).

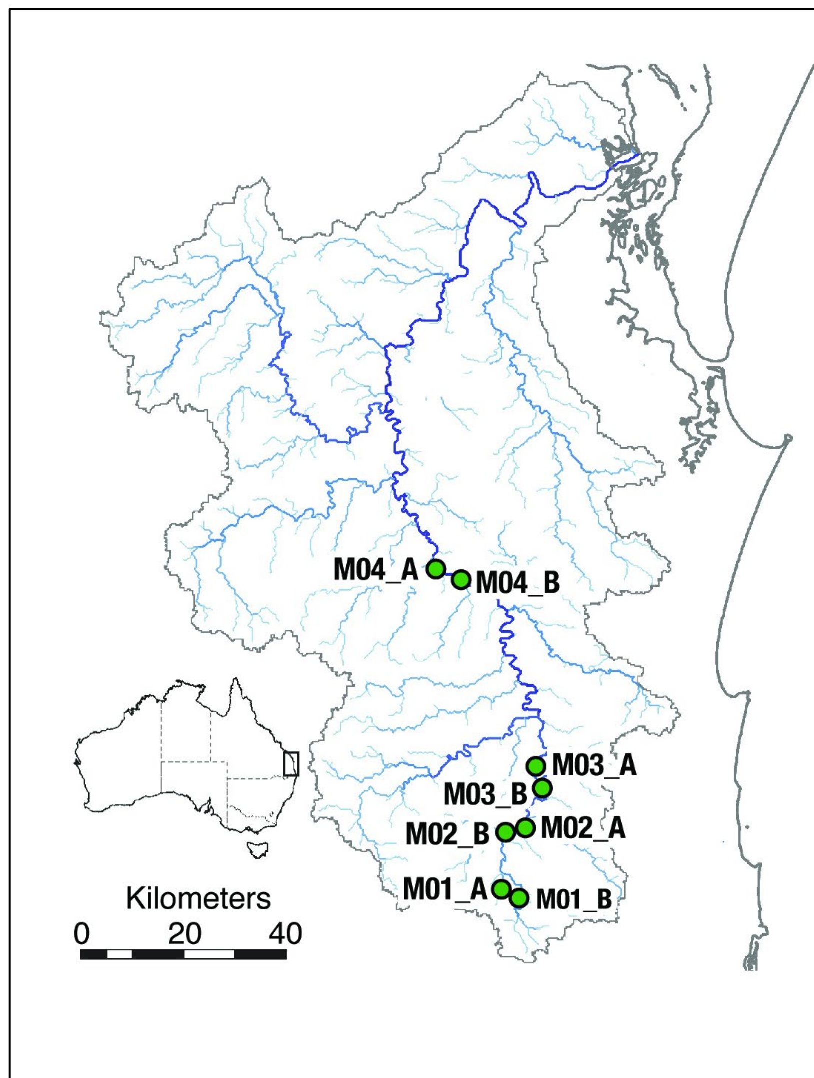

Figure 4.1: The location of sampling sites (green circles) within the Mary River catchment, eastern Australia

Extractions were performed in a dedicated DNA extraction facility that is geographically separated from where PCR reactions were undertaken to avoid contamination. 30 µL of Proteinase K (20 mg/ml, Qiagen, Valencia, CA) was added to sample vials that already contained the filters and SET buffer (0.5% SDS), as per field protocol. Sample vials were incubated with rotation overnight at 56°C. 200 µL of this solution was pipetted into a sterile tube (Eppendorf) with an equal volume of Phenol:Chloroform:Isoamyl Alcohol. The following steps followed standard procedures for Phenol:Chloroform extraction (Sambrook and Russell, 2006). 200 µL of supernatant was mixed with 400 µL of NTC buffer (Machery-Nagel). The remaining extraction was performed with the NucleoSpin® Gel and PCR clean-up Kit (Machery-Nagel) following manufacturer's protocol, with the exception of one extra wash with NT3 buffer and final elution using 30 µL of 0.01% Triton X100. Extracts were stored at -20°C.

Standard PCR using metabarcoding primers targeting bony fish was conducted following contamination control protocols suitable for ancient DNA analysis (Paabo et al., 2004). PCR conditions are described in Appendix 4.1. Illumina adapters were added following the Illumina NEXTERA 16S amplicon sequencing protocol (Illumina, San Diego, CA), with precise conditions described in Appendix 4.2.

Products from positive reactions and negative controls for each of the primer sets were pooled. Pools for each primer set were cleaned separately with AMPure XP beads (Nextera XT DNA 150 kit, Illumina, Inc., San Diego, CA, USA) following the manufacturer's protocol, except that a ratio of bead to product volume of 0.9:1 was used. The size of the clean products was determined on a Bioanalyzer DNA 1000 chip (Agilent Technologies, Santa Clara, CA, USA) and quantified with a Qubit (1.0) fluorometer (Agilent) and qPCR using the NEBNext Library Quant Kit for Illumina (New England Biolabs, Inc., Ipswich, MA, USA), following manufacturer's instructions. Based

on qPCR and bioanalyzer quantitation results, an equimolar amount of the sample pool for each primer set was added to a final pool combining the product from all three primer sets. This final pool was diluted with molecular grade water to 4nM for paired-end sequencing on an Illumina MiSeq at the Griffith University DNA Sequencing Facility, Brisbane, Australia following the manufacturer's protocols (Illumina, Inc., San Diego, CA, USA).

4.3.3 Bioinformatics Methods

Sequence reads were demultiplexed by sample using the MiSeq native bcl2fastq2 command. Reads were merged using PEAR v0.9.10 (Zhang et al., 2013), then quality filtered, dereplicated, checked for chimeras and clustered at 97% using UPARSE version 8.0.1623 (Edgar, 2013). Taxonomic assignment was performed using a naïve Bayesian classifier (Wang et al., 2007) against a custom database. The latter database was constructed by *in silico* PCR (Bellemain et al., 2010, Ficetola et al., 2010) using the primers from the present study (Table 4.1) to amplify homologous DNA regions from the NCBI nt database (accessed June 2016) allowing up to three base pair mismatches per primer.

Table 4.1: List of primer names, associated sequences, melting temperatures used for study, mean length of target region plus primers when aligned with various reference genomes for fishes from GenBank (NCBI), and references for two primers sets taken from the literature

Primer Name	Primer Sequence	T_m (°C)	Mean length (bp)	Reference
Teleo_F	ACACCGCCCGTCACTCT	55	101	(Valentini et al., 2016)
Teleo_R	CTTCCGGTACACTTACCATG			
MiFish_F	GTCGGTAAAACTCGTGCCAGC	66	220	(Miya et al., 2015)
MiFish_R	CATAGTGGGGTATCTAATCCCAGTTTG			
Ostei12S_F	GGATTAGATACCCCACTATG	55	104	This study
Ostei12S_R	GYCAAGTCCTTTGGGTTT			

4.3.4 Species distribution modeling

Species distribution models developed by Rose et al. (2016) were used to predict the probability of occurrence for different genera at each sampling location, based on their environmental characteristics. Predictions were derived from an ensemble of five commonly used single-species models fitted using the BIOMOD package (Thuiller et al., 2009) in R (version 2.1.3, Foundation for Statistical Computing: Vienna, Austria). The models used were: generalized linear models (GLM), random forest (RF), boosted regression trees (GBM), artificial neural networks (ANN) and multivariate adaptive regression splines (MARS). Parameter settings used for each model were the default in the BIOMOD package. Models were trained and validated using fish assemblage data collected from 128 least disturbed sites (mean stream length of 98m) in coastal rivers of sub-tropical eastern Australia extending from the Mary River Basin in south-eastern Queensland, to the Clarence River Basin in north-eastern New South Wales. Sites were sampled in the post-wet season (autumn/winter) of 2013 using a standardized single pass backpack electrofishing protocol, which yields reliable estimates of species presence-absence at each site. Fish catch data were converted to species presence-absence, and alien and estuarine vagrant species were removed prior to analyses. Species occurring at only one site in the training dataset were also removed, as these could not be modelled accurately. The final training dataset consisted of 25 native fish species from 19 genera. Twenty minimally-redundant environmental variables describing climate, catchment topography, geology and runoff characteristics were used as candidate predictor variables in the models. Models were trained using 80% of the dataset (n=103 sites) and k-fold cross-validated (k = 5) using the remaining 20% (n=25 sites). This process was repeated 10 times for each modeling algorithm, using a randomized 80/20% allocation of sites each time, resulting in 50 candidate models per species for ensemble modeling. Only cross-validated models were retained for the

ensemble model, that is, the full model was not considered. The true skill statistic (TSS) was used as the criterion for model retention (Allouche et al., 2006). Models were retained if the TSS>0.8, except where no candidate models met this criterion, in which case, TSS> 0.7 was used. The retained models were combined into a single ensemble model for each species by calculating the arithmetic mean of prediction probabilities. This method was used as it has been shown to provide more robust predictions than single models and most other consensus techniques (Marmion et al., 2009). The ensemble models were shown by Rose et al. (2016) to be more accurate and precise than other modeling methods. The ensemble models were used to predict the probability of occurrence of each species at each of the eight locations sampled for eDNA based on their environmental characteristics. The predicted probability of occurrence of each genus at each location was estimated simply as the maximum of the individual species' prediction probabilities within each respective genus. The model results were used to help determine plausibility of eDNA metabarcoding results for taxa not detected by seining.

4.3.5 Data analysis

Several samples produced insufficient read counts or quality of reads and were withheld from further analysis. Rarefaction of datasets was conducted, following the findings of Deagle et al. (2018). Rarefaction addresses the issue that the higher the number of sequences sampled, the higher the number of OTUs that will be identified; therefore it enables calculation of richness relative to the number of sequences sampled. This is done by randomly re-sampling a portion of sequences from a large pool many times and calculating the average richness (Heck et al., 1975). For hypothesis 1 (multiple clade-specific primers will overcome sequencing drop-out and amplification biases), analysis was done at the genus level primarily because the reference database was not detailed enough at the species level and lacked

representation of many Australian freshwater fish species. Also, the short fragments of DNA length targeted for eDNA metabarcoding do not always have species-level resolution. For hypothesis 2 (PCR and sequencing biases occur equally across samples), OTUs were used for analysis, therefore the incomplete database was not an issue.

The first hypothesis, that the use of multiple universal primers will overcome drop-out from sequencing and amplification biases to provide more accurate taxon composition estimates, was tested. A comparison among the number of unique OTUs and genera identified by each primer set on its own was made. The results from each primer set were also compared to the total number of OTUs and genera identified when results from all primer sets were combined. To calculate genus-level richness for each marker, sequences that could only be assigned down to the order or family level were counted once and scored as a single genus to present the most conservative results. Therefore, genus-level richness data presented in this study represent minimum estimates, and may underestimate the total number of genera. To determine the plausibility of metabarcoding results, these taxon composition estimates were compared to alternative sources of composition data: seining observations, species distribution modeling, and a reference list of fishes known to occur in the Mary River catchment based on occurrence data reported in a large number of studies using a variety of conventional sampling methods (Pusey et al., 2004).

The second hypothesis, that primer biases occur equally across each water sample, predicts beta diversity will be consistent among primer sets. It should demonstrate that all primer sets concordantly index diversity across samples. This hypothesis was tested by comparing between-sample diversity estimates (Table 4.2)

produced by each primer set with Mantel tests (Table 4.3) and heat-mapping (Figure 4.2). For the Mantel tests two matrices were produced for each primer set: (1) Using a Bray-Curtis index on the sample raw reads x taxon data, and (2) using a Jaccard's index based on presence-absence of taxon. Dissimilarity matrices were compared using Mantel tests based on Pearson's product-moment correlation with 999 permutations in the software R version 3.3.1 (R Core Team, 2016) with the package *vegan* (Oksanen et al., 2013). For the heatmap the proportion of reads (%) for each marker that were assigned to specific orders, families and genera were compared (Figure 4.2). Cluster dendrograms (UPARSE in the UCLUST package (Edgar, 2013)) were constructed using a Bray-Curtis distance matrix. This was then used to test whether taxon composition estimated for each site, based on the proportion of reads for each of the three markers, were grouped by sample (Figure 4.2). Several samples were left out of analysis because of an insufficient number or quality of reads.

4.4 Results

The number of OTUs identified using each marker in rarefied datasets at a 97% clustering threshold was 71, 24, and 35 for the Teleo, Ostei12S, and MiFish markers respectively (Table 4.2). The Teleo primers showed minor variation in OTU count at higher clustering thresholds, but this number declined sharply at a 95% threshold. The Teleo markers produced the highest number of OTUs across all clustering thresholds. The Ostei12S markers produced the most consistent number of OTUs at different clustering thresholds. The number of OTUs produced by the MiFish primers declined gradually along with a decreasing clustering threshold. The total number of genera identified by each marker in rarefied datasets was at least 12, 10, and 8 for the Teleo, Ostei12S, and MiFish primers respectively (Table 4.2) for all sites combined. In the case of results that could be assigned to the genus level, at least 15

unique genera were correctly identified by all three markers, across all sites (Figure 4.2).

Table 4.2: The number of OTUs produced and genera identified in rarefied datasets from eDNA metabarcoding with three universal fish markers; summed across all sites, at a 97% clustering threshold. Orders or families that could not be assigned to the genus level are counted once, and scored as a single genus. Counts only include taxa with a proportion of sequence reads ≥5%.

Primer Set	OTU Count	Genus Richness
Teleo	71	12
Ostei12S	24	10
MiFish	35	8

A dendrogram indicated that taxon composition and diversity estimates did not group either by sample, or by the three sets of markers (Figure 4.2). For example, the dendrogram clustered five samples (M01A, M02B, M03A, M03B, and M04A) in which a high proportion of reads from all three markers were assigned to *Hypseleotris*. However, for sample M03A, only the *Hypseleotris* sequences detected by the MiFish markers were included in this cluster. The Teleo primer set did not detect the presence of *Hypseleotris* in sample M03A. This genus was detected in sample M03A by the Ostei12S markers, but these reads were clustered separately in the dendrogram. This distinct cluster of three samples covered at least six genera, but only contained reads from the Ostei12S primers. Across all samples and primer sets, two genera (*Hypseleotris* and *Nematalosa*) were identified in a higher proportion of reads than any others.

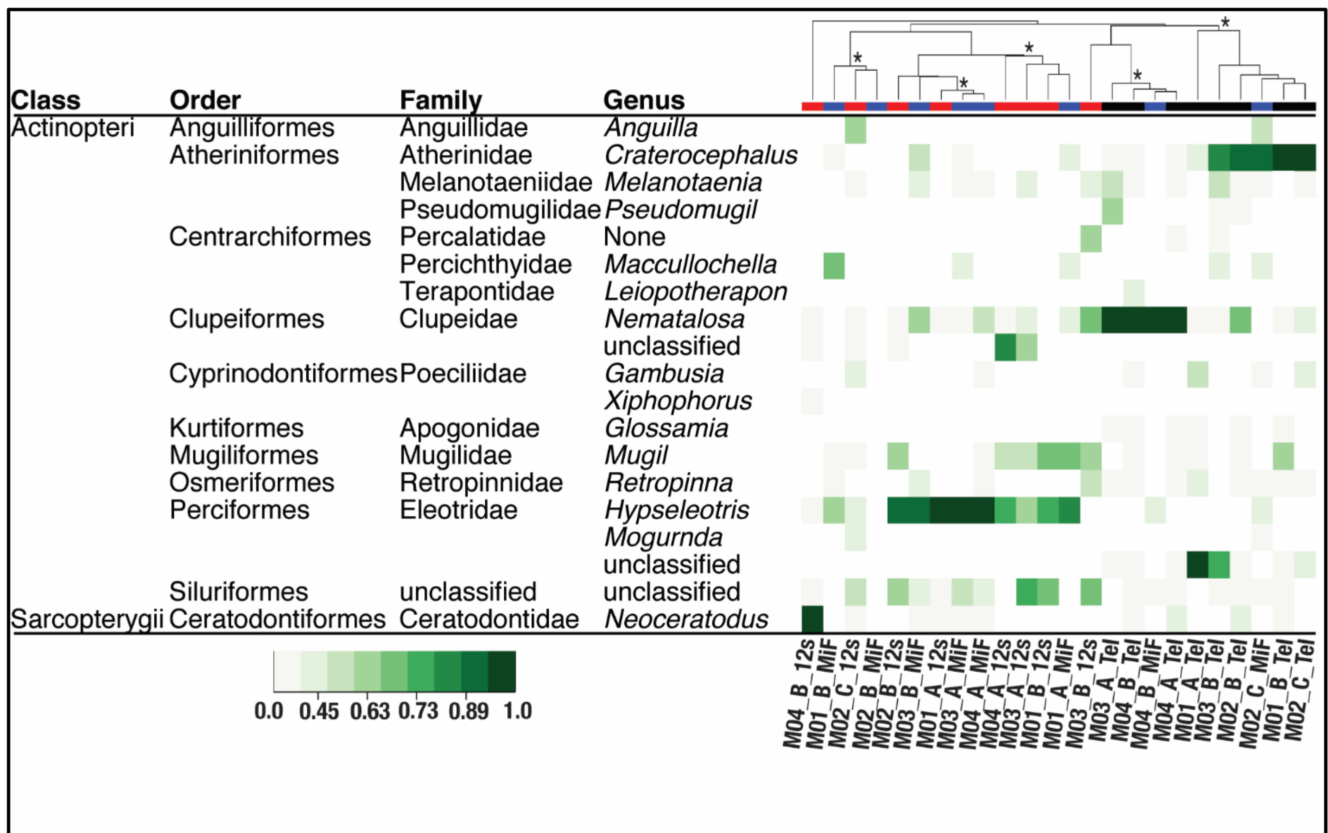

Figure 4.2: Taxon composition for assigned sequences in each sample based on eDNA metabarcoding with three generic fish markers (Sample ID's indicate site location with underscore to indicate primer name, with 12s= Ostei12s, Mif = MiFish, and Tel = Teleo). Green shading intensity shows relative proportion of total reads assigned to each taxon by each marker. To help visualization of less abundant taxa the proportions have been square root transformed (note non-linear scale). Dendrogram is based on average clustering of a Bray-Curtis dissimilarity matrix derived from the proportion of fish genera amplified at each site. Color bar in first row shows primer identity (black = Teleo, red = Ostei12S, blue = MiFish). Asterisks on the dendrogram are at the nodes of clusters that are statistically more supported than chance (alpha = 95%, SIMPROF). Note that sequence libraries do not cluster by primer set, which would be predicted if primer biases fully determined composition. However, samples do not cluster by the eDNA extract that they were amplified from, which would be predicted if there were no primer biases or sequencing artifacts. Instead, while the composition of DNA sequences does supervene on the true assemblage, and therefore represents it, it represents it somewhat stochastically.

Mantel tests showed positive, though mostly weak, correlations between dissimilarity matrices produced using the different markers, both within and among sampling sites. Pearson's *r* values ranged from a minimum of 0.137 (Teleo x MiFish with Bray-Curtis) (Table 4.3a.) to a maximum of 0.782 (Teleo x Ostei12S with Jaccard's) (Table 4.3b.). The best correlations were between the Teleo and Osteo12s markers

using Jaccard's Index (p =0.001) (Table 4.3b). All correlations with the MiFish markers were poor (all <0.423) (Table 4.3).

Table 4.3: Mantel test statistics

4.3a. Pearson's correlation coefficients comparing Bray-Curtis indices on raw data, with rarefication at 2500-reads, and excluding OTUs that were unidentified (support <0.5 for matches with GenBank references).

	Pearson's *r*	
	Ostei12S	Teleo
Teleo	0.655	
MiFish	0.167	0.137

4.3b. Pearson's correlation coefficients comparing Jaccard's indices, with rarefication at 2500-read depth, and excluding OTUs with support <0.5 for matches with GenBank references. The only significant correlation for all Mantel tests was found for comparison of matrices for the Teleo and Ostei12S primers under the above-mentioned filtering protocols (*p = 0.001).

	Pearson's *r*	
	Ostei12S	Teleo
Teleo	0.782*	
MiFish	0.337	0.423

Estimates of generic composition based on eDNA analysis, seining data, and probability of occurrence from species distribution modeling showed varying levels of agreement across all sites (Figure 4.3). Eight genera in total were collected with seining. On their own, 50%, 37.5%, and 37.5% of these genera were identified by the Teleo, Ostei12S, and MiFish primer sets, respectively. In tandem, these markers identified 62.6% of genera observed with seining, failing to identify three (*Ambassis, Craterocephalus, and Oreochromis*). For genera included in species distribution modeling, 11 were predicted to be "highly likely" to occur in the sampling area (associated with an occurrence probability ≥0.6) (Figure 4.3) On their own, 27.3%,

36.4%, and 36.4% of these genera were detected by the Teleo, Ostei12S, and MiFish primer sets, respectively. The combined results from all markers identified 54.5% of genera with high probability of occurrence. One genus of gudgeon (*Kimberleyeleotris*) that was detected by metabarcoding is not known to occur in the sampling area.

Different sources of taxonomic composition (assemblage) data showed high agreement for some genera, but returned inconsistent results for others (Figure 4.3). Results were congruent for the gudgeon genus *Hypseleotris*, the blue-eye *Pseudomugil*, the smelt *Retropinna*, and genera in the Rainbowfish family (*Melanotaeniidae)*. The rainbowfish genus *Melanotaenia* is known to occur in the Mary River (Pusey et al., 2004), was identified with eDNA, observed with seining, and models suggested that they are likely to have high probability of occurrence. The rainbowfish *Rhadinocentrus* was neither detected with eDNA methods nor seining, and had a low occurrence probability. Some genera of gudgeon (*Mogurnda*) and eel (*Anguilla*) were not observed with seining; yet they are known to occur in the river (Pusey et al., 2004), were identified with eDNA methods, and models suggested that they likely have high occurrence probabilities. Sources of taxonomic composition data were inconsistent for the grunter genus *Leiopotherapon* and mullet genus *Mugil*. While these genera were identified with metabarcoding and are known to occur in the river; the probability of occurrence was "moderate" ($0.3 < p \leq 0.5$) and they were not detected by seining. Mary River cod (*Maccullochella mariensis*), platy (*Xiphophorus maculatus*), an unknown genus of catfish (Order Siluriformes), and Australian lungfish were identified in eDNA results, but not detected with seining or predicted to occur (Figure 3). In contrast, the siluriforme genus of *Tandanus* was predicted to occur with high probability of occurrence, but was not identified with eDNA. The genus *Oreochromis* was detected with seining, but not by any of the

generic fish markers in sequencing results. This indicates either a failure of the primers to amplify this genus or dropout of the genus in sequencing.

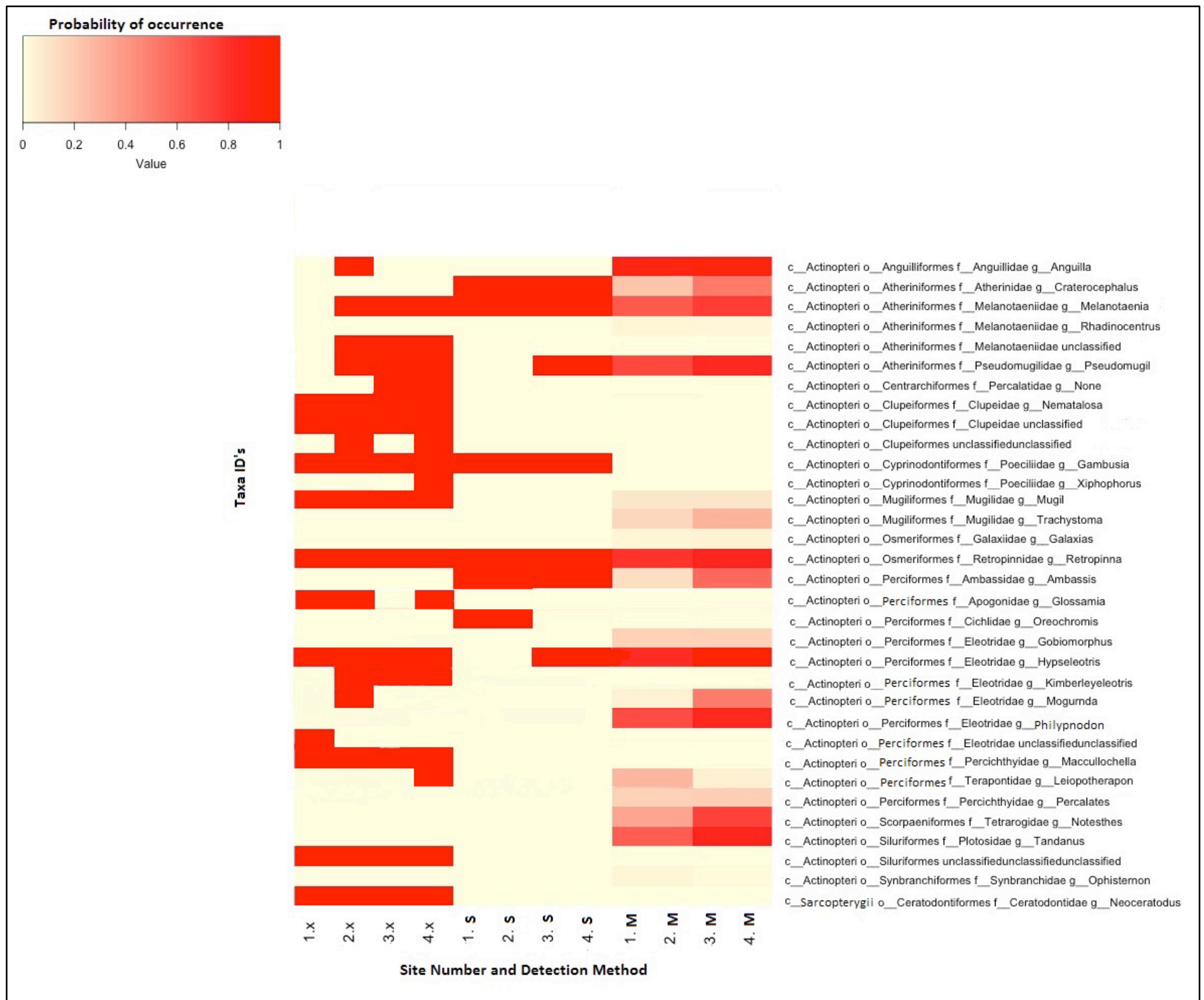

Figure 4.3: Comparison of taxon composition estimates at four sampling sites produced by all three generic primers with eDNA metabarcoding (1-4.x), conventional sampling (1-4.S), and species distribution modeling (1-4.M). Color key indicates probability of occurrence for modeling results, with 100% probability shading for metabarcoding and seining results where taxa IDs were certain, and buff shading where taxa were not detected or observed.

4.5 Discussion

The findings of this study are relevant to current trends in aquatic research and conservation. A recent review found that 46% of twenty-eight publications using eDNA to study occupancy in aquatic ecosystems included fish (Roussel et al., 2015).

Metabarcoding studies that use universal primers to assess taxonomic composition have mostly been conducted in the field of bacterial ecology (Bryant et al., 2012), and more recently and to a much lesser extent used to assess the diversity of fungi (Tedersoo et al., 2014), protists (Pawlowski et al., 2014), and plants (Pedersen et al., 2013). Several recent publications that used eDNA to detect taxa have also used it to estimate richness and diversity of aquatic macroorganisms (Valentini et al., 2016, Olds et al., 2016), and the number of such studies and recommendations for procedures is rapidly growing (Rees et al., 2014, Evans and Lamberti, 2018)

The broad goal of this study was to understand how reliably eDNA sequencing with different generic fish primer sets can represent true composition, richness and beta diversity. If the generic markers all returned similar results (as anticipated given the expectation of primer bias being equally applied across samples), then abundance of reads assigned to a genera should have been clustered by water sample. This was not the case. The results showed there was no discernible clustering of results by water samples (Figure 4.2). The clusters which were apparent were not grouped by the three primer sets either (Figure 4.2), which would be expected if the markers were giving completely divergent results. These findings indicate that there are PCR biases and that different primer sets have different biases; therefore, the choice of "universal" marker influences richness and diversity estimates and biases do not occur equally across samples. It is possible that the potential for variability throughout the molecular workflow could influence results. For example, when quantities as small as 1µL of eDNA extract are added to PCR reactions, there could be a small variation in the quantity of molecules added to each tube. Using the same pipette to set up all PCR reactions will reduce the potential for this type of variability.

Metabarcoding is increasingly being applied in a conservation context, and the results reported as beta diversity indices (Chain et al., 2016, Evans et al., 2016). If molecular data are used to inform management, conservation, and development decisions, the ability to quantify any bias is essential for predicting the effectiveness of policies based on them. For areas lacking baseline data or where survey with conventional methods is problematic, the current limitations of molecular methods are likely outweighed by the conservation benefits of extracting data from inaccessible and unexplored habitats.

4.5.1 Hypothesis 1: Use of multiple universal primer sets will provide more accurate richness estimates

Higher and more accurate richness estimates were obtained by using multiple sets of universal fish primers than by using a single marker alone. The Teleo primers produced the highest genus richness estimates, indicating that they amplified a higher number of unique sequences; yet it is unclear whether these sequences all represent true species. The high number of unique sequences could be the result of PCR artifacts, or the species these sequences come from may not be well-represented in the GenBank reference database. Taxa collected via seine-netting validated most of the eDNA detections. As some genera are not vulnerable to capture by seine-netting, species distribution modeling and the reference list of species known to occur in the river is useful for determining levels of confidence in eDNA detections of taxa not observed with seining.

Overall, at the genus and species levels, the use of multiple primer sets helped fill gaps in taxon composition estimates produced by a single marker alone. Taxonomic coverage and depth varied among primer sets, but there were several families (Ambassidae, Atherinidae, Cichlidae, Hemiramphidae, and Retropinnidae) that were not detected by any of the markers after quality filtering, though several

species from these families are known to occur in the river. The Teleo primers were the only markers that did not identify either of the species of *Anguilla* that are known to occur in the river, but were also the only markers to accurately detect *Pseudomugil signifier* and *Glossamia aprion*. Anguillidae is highly divergent from many other families of bony fishes. This fact complicates the design of a single universal set of markers that can efficiently amplify this family simultaneously with more recently evolved families. Eels typically have low susceptibility to capture by seines, but modeling results indicate moderate confidence in the metabarcoding detections of *Anguilla*.

The use of multiple primer sets increased the potential for metabarcoding to detect the same taxa observed with seine netting. Metabarcoding detected at least eight more genera than seine netting. The likely presence of these genera was confirmed with species distribution modeling, and a reference list of fish species occurrence from data reported in a large number of studies using a variety of conventional sampling methods (Pusey et al., 2004). The Ostei12S and Teleo primers identified the same number of incorrect species (according to Pusey et al. 2004), but the erroneous species they amplified were different except for one. *Leiopotherapon plumbeus* was incorrectly identified; though *Leiopotherapon unicolor* is known to occur in the Mary River. Initially, some sequences were assigned to several non-fish vertebrates (Orders: Anseriformes, Anura, Bovidae, Chiroptera, Monotremata, Phyllodactylidae, Primates, Squaliformes, and Testudinidae), though filtering removed them prior to final analysis. The primer sets amplified different non-fish taxa except for Bovidae, which was amplified by both MiFish and Ostei12S.

Primers that will accurately resolve to the species level for sympatric species are difficult to design for the short fragments typical of eDNA, when attempting to

target numerous species from multiple taxonomic orders. Traditional barcodes are usually not suitable because of their length (Hajibabaei et al., 2006). Some studies address this problem, and PCR bias in general, by using multiple primers that are species-specific (Wilcox et al., 2013, Fukumoto et al., 2015) or several clade-specific primers in tandem (Evans et al., 2016, Olds et al., 2016). If taxonomic analysis is required, richness estimates based on metabarcoding are limited by the availability of reference sequences. Reference sequences for genomic regions appropriate for eDNA analyses are currently not publicly available for many fishes. Increased investment in this resource and the taxonomic expertise required to obtain it is warranted, given the rapid uptake of molecular survey methods. The results of this study show that it is best to use several generic markers simultaneously.

4.5.2 Hypothesis 2: Biases occur equally across samples; therefore beta diversity will be consistent among primer sets

This hypothesis was assessed by converting the samples to beta diversity matrices and evaluating whether the same samples always cluster together, regardless of primer choice. After comparing the diversity matrices with Mantel tests, it was found that this was not the case and results showed that primer choice influenced results. While the interpretation of metabarcoding results is influenced by the algorithms and methods used for bioinformatics and analyses (Bachy et al., 2013, Flynn et al., 2015), it is unlikely that the variation found in results from this chapter is due to bioinformatic analyses in this case. Initial results showed surprisingly high variation in between-sample diversity estimates among the three markers, which could be caused by uneven or low sequencing depths (Smith and Peay, 2014). The final results presented here support the interpretation that each generic fish marker produced biased estimates of fish richness.

The large dissimilarity in taxonomic composition shown between samples (Figure 4.2) that was found using three group-specific fish primer sets can be attributed to several factors. Possible causes include environmental variation at sampling sites, bias from molecular workflows, differences in the target lengths of markers, preferential amplification by primers, and threshold selection for bioinformatics analysis. One explanation could be that the Ostei12S primers were designed to amplify Australian lungfish (class Sarcopterygii) along with other bony fishes, while the Teleo primers were designed for an infraclass (Teleostei) of Actinopterygii. Frogs and other tetrapods evolved from Sarcopterygii, while the largest group of bony fishes diverged from Actinopterygii. Despite this, *in silico* validation showed that the Teleo markers did align with Australian lungfish reference genomes downloaded from GenBank (barring one G/A mismatch in reverse end primer), so genetic divergence and associated efficacy of marker is likely not the cause of observed variation. The inconsistency in diversity estimates produced between the Teleo and Ostei12S primers may be the result of PCR artifacts.

When samples include a complex mixture of templates from many species, and generic markers are used to amplify specific genes simultaneously, it is known that certain templates may be stochastically underamplified (Polz and Cavanaugh, 1998). A large number of PCR cycles can exacerbate this problem of stochasticity. In the present study a large number of PCR cycles (50 total) were used for amplification and library building; however, this is the same number used in the study where the Teleo primers were originally published (Valentini et al., 2016). Additionally, a large number of cycles is typical for analysis of degraded and ancient DNA samples (Taberlet et al., 1996, Hofreiter et al., 2001). The analysis presented here showed the highest proportion of reads from the MiFish primer set were assigned to *Hypseleotris*

(77.97%), which could indicate preferential amplification of members of this genus. Compared to the Ostei12S reads, a relatively high proportion of sequences from the Teleo and MiFish markers were assigned to *Nematalosa*. This supports the interpretation that markers are biased towards different taxa, despite systematic attempts to design primers that amplify all bony fishes.

The dendrogram (Figure 4.2) indicated a higher dissimilarity in results produced by the MiFish primers than the dissimilarity between results from the Teleo and Ostei12S markers. This higher dissimilarity may be attributed to the fact that the MiFish primers target a longer fragment (172bp) than the Teleo (64bp) or Ostei12S (66bp) primers. DNA typically degrades quickly in water after being shed, but degradation rates vary substantially depending on environmental factors (Strickler et al., 2015). Because the MiFish primers target longer fragments of eDNA, these results could be skewed towards preferential detection of higher-quality molecules. Moreover, eDNA shedding and degradation rates are variable among species (Evans et al., 2016) and life-stages (Maruyama et al., 2014). A marker which targets a longer fragment may preferentially amplify taxa that were present more recently, present in higher densities, or that shed DNA at higher rates.

4.5.3 Conservation Implications

The results of this chapter showed that when multiple universal fish markers were used, a higher level of generic richness was detected, and a higher level of richness was detected with a combination of molecular and conventional methods. The results indicate that the choice of "universal" marker is likely to skew richness and diversity estimates. If the goal of an eDNA metabarcoding study is to accurately assess differences in diversity among samples, the use of a single universal marker may not be sufficient. Beta diversity indices are a ubiquitous measure of biodiversity

used in eDNA studies, yet the results presented here and elsewhere indicate that beta diversity indices based on metabarcoding are likely to be skewed (Dickie, 2010, Deagle et al., 2013, Ficetola et al., 2015). The use of multiple primer sets, control datasets, *in silico* PCR, and/or occupancy models to quantify detection probabilities and inform study design from the first stages is highly advisable. Sequencing depth may have greater influence on the sensitivity of eDNA metabarcoding results than the number of PCR replicates (Smith and Peay, 2014). For any random water sample, the eDNA of rare species is likely to be present in low concentrations relative to abundant taxa. With low or uneven sequencing depths, the amplicons from species that are present in low density may not be detected. Quantifying the accuracy and sensitivity of a sampling method is a prerequisite for determining appropriate levels of confidence in interpretation of results. If diversity indices based on eDNA metabarcoding are used for conservation planning, resource management, or infrastructure development decisions (i.e. location of impoundments), the level of confidence in these data will influence the outcomes of cost-benefit analyses. It is recommended that data be interpreted in a format that is relevant for policy-makers. Appendix 3 of this chapter provides an example of how these data can be presented in a format that is relevant and accessible for policy and management contexts.

For comprehensive aquatic biodiversity assessment using conventional methods, it is often necessary to use multiple sampling gear types. This is the case for large tropical rivers with high species diversity and heterogeneous habitat. These systems arguably have the greatest need for surveys, as currently available diversity and distribution data are typically limited (Dudgeon, 2000, Winemiller et al., 2016). For many regions, this is due to the logistical challenges encountered when sampling with conventional methods (Bellemain et al., 2016). Tropical rivers are increasingly

subject to infrastructure development and resource extraction for human uses (Dudgeon, 2000, Winemiller et al., 2016), which are likely to cause major shifts in species diversity (Sodhi et al., 2010). The majority of diversity data for such systems come from fishery-dependent catch surveys, fisher interviews, and fisher logbooks (Silvano et al., 2008, de Graaf et al., 2011). While these data are particularly useful, they are not likely to represent true species diversity. Fishers often use selective gears and avoid reporting capture of rare or threatened species. For this study, discrepancy in the number of taxa detected with eDNA metabarcoding compared to seining is expected, given the selective nature of seine nets. Electrofishing is considered the most sensitive conventional sampling method, and when used with a multiple-pass sampling methodology has been shown to accurately assess species composition and abundance in systems with moderate diversity (Kennard et al., 2006). The primary benefits of eDNA sampling over electrofishing include its ease of application in habitats unsuitable for electrofishing, its non-destructive nature, potentially higher detection probabilities, reduced need for taxonomic expertise, and lower cost (Evans et al., 2017b). However, in many cases application of conventional survey methods like electrofishing can provide a more comprehensive ecological perspective than eDNA methods currently are able to on their own. No single sampling method can overcome the challenge of false-negative detections for rare species, but guidance for quantifying eDNA detection probabilities is accumulating in the literature (Schmidt et al., 2013, Ficetola et al., 2015, Hunter et al., 2017).

Conservation is often the ultimate goal of biodiversity assessment, and the priority species for conservation interventions are typically present at low densities. It is important to recognize that eDNA metabarcoding may not detect rare species at the same rates as highly abundant species, and a single universal primer is unlikely to

provide comprehensive biodiversity data. At this stage, eDNA metabarcoding data are likely unsuitable for abundance-based estimates of diversity, and are more appropriate for presence-absence based measures of diversity such as richness (Roussel et al., 2015). Despite current limitations, the vast data that are being made accessible by eDNA metabarcoding represent a step change for aquatic species research, and are of especially high value for conservation planning. Conservation efforts have a higher probability of success with increased spatiotemporal resolution of diversity data. eDNA methods can provide such data when applied appropriately.

4.6 Conclusions

Despite rapid growth in popularity, the choice to apply eDNA metabarcoding should be informed by the unique characteristics of the study system, species, and questions being addressed. The concept of a single, comprehensive, and rapid aquatic diversity survey approach is tempting, but results presented here indicate that eDNA metabarcoding using a single "universal" fish marker does not provide complete richness or accurate β diversity estimates. This study highlights the benefits of metabarcoding as a complement to conventional sampling and species distribution modeling, while emphasizing the importance of thoroughly assessing the influence of molecular markers on results. This conclusion is supported by the findings of other studies, which showed that PCR efficiency is highly species-specific (Polz and Cavanaugh, 1998, Elbrecht and Leese, 2015). To quantify PCR bias, concurrent sequencing of control data sets (a mixture of templates in known quantities, extracted from sources with verified taxonomy) along with eDNA samples is recommended (Sommeria-Klein et al., 2016). The results of this study add further support to such a recommendation. Given the relatively high decline of freshwater species compared to terrestrial taxa (Ricciardi and Rasmussen, 1999) and broad loss

of the planet's freshwater biodiversity (Dudgeon et al., 2006), the application of molecular tools should not be postponed because of existing limitations. However, further work is needed to assess limitations, and a thorough review of limitations should be included in the design and reporting of research.

Animal Ethics Statement:

No animals were collected for this study. All species presence data gathered by seine-netting was collected for a separate study. The Griffith University Animal Ethics Committee approval number from for the separate study is ENV/12/14/AEC.

CHAPTER 5: LOCATION OF WATER SAMPLING POINT WITHIN A RIVER CHANNEL INFLUENCES EDNA METABARCODING RESULTS: THE CHALLENGE OF RELIABLY SAMPLING EDNA FROM RIVERS

5.1 Summary

Published studies using Environmental DNA (eDNA) methods in rivers typically utilize samples that are collected at the river bank, and the type of hydraulic unit (i.e. riffle, run, or pool) sampled is rarely reported. In addition, such studies implicitly presume that a water sample collected from the bank at any time of day will contain the same eDNA composition and quantities. However, given the heterogeneous nature of rivers and variation in water velocity, temperature, and UV light exposure, this assumption must be tested. Water velocity influences the transport distance, resuspension, and concentration of eDNA in rivers. This is the first example of a study that compared samples collected at the river bank with samples collected in other portions of the channel, and compares results among hydraulic unit types.

The following study tests the assumption that samples collected at the river bank are representative of eDNA from the river. To do this, replicate sampling was performed at two additional points across the river channel. Additionally, the influence of the time of sample collection and the hydraulic unit type was tested by collecting samples in a pool and in a riffle immediately downstream. A cross-section survey was conducted using Acoustic Doppler Velocimetry (ADV) to measure the velocity at different depths and produce isovel profiles. Surface water samples were collected in triplicate for each sampling location along transects, three times within a single day. eDNA was amplified with PCR using both a universal bony fish primer "Teleo" (Valentini et al., 2016) and a universal 18S eukaryote primer (Jarman et al., 2013), and subjected to high-throughput sequencing.

The results of this study showed that water velocity varied substantially both at the surface and at different depths within a full cross-section. Within each transect there was a high degree of variation in the number of OTUs identified among sampling points based on eDNA metabarcoding results. Comparisons between different hydraulic units (pool and riffle) also showed variability in results. Findings indicate that there is a high degree of variation in samples collected in relatively close proximity to one another and, for eukaryotes, this variation was significantly related to complex interactions between time of sampling, proximity to river bank, and hydraulic unit type. Results using fish primers also appeared idiosyncratic, but were not found to be significantly related to environmental factors tested. These findings have important implications for eDNA metabarcoding studies which aim to assess taxa richness or assemblage for a given river reach. A number of samples taken from different points within the sampling unit are likely required for accurate assessments. Further work is required to test appropriate sampling stratification along ecologically relevant environmental gradients in rivers.

The supposed advantages of eDNA methods over conventional methods include increased sensitivity and ability to gain information on species composition or presence without *a priori* knowledge of such information. Studies which rely on both conventional methods and eDNA are typically aimed at validating eDNA results. For the method to achieve its potential utility through broad applications, the reliability of results without validation from conventional sampling needs to be tested. There is great conservation potential for application of the methods, especially for environments with high diversity, high number of threatened species, and with little prior knowledge of distributions, such as large tropical rivers in developing countries or geographically isolated regions. However, very limited work using eDNA has been conducted in such environments. Further work is needed to understand how the heterogeneity of rivers

influences the potential to collect representative water samples, especially in natural environments and warmer climates.

5.2 Introduction

Environmental DNA (eDNA) methods are increasingly being applied for research and management in aquatic ecosystems (Rees et al., 2014, Stein et al., 2014, Roussel et al., 2015, Thomsen and Willerslev, 2015). The potential for improved conservation outcomes from such data is high; however, more work is needed to establish best practices for sampling in a variety of environments. Rivers are challenging environments to sample for eDNA. Accurate determination of detection probabilities and interpretation of results are complicated by the constant flows and high longitudinal and latitudinal variation in environmental conditions. These factors likely influence the transport, retention, and degradation rates of eDNA (Barnes and Turner, 2015, Jerde et al., 2016, Shogren et al., 2017). Water samples for eDNA analysis are typically collected from the surface near the river bank at a single time and point location (Laramie et al., 2015). However, this information and the type of hydraulic unit (i.e. pool, riffle, or run) are not reported in most publications. The above sampling strategy implicitly presumes that eDNA is homogenized in the river channel and that a single point location will contain eDNA concentrations and mixtures that are representative of the eDNA of all organisms present at the site and some distance upstream. Given knowledge of organic particle transport from other well-established fields, such as water quality monitoring and hydrogeomorphology (Newbold et al., 2005, Aubeneau et al., 2014), this assumption is likely flawed.

eDNA can degrade rapidly in water once shed, with an estimated half-life of ~6 hours under certain conditions (Barnes et al., 2014). UV light is highly influential in the degradation of DNA. UV light penetration varies across a river channel and throughout

the day (Pilliod et al., 2014). In addition, river discharge levels have been shown to influence eDNA detection probabilities (Shogren et al., 2017). The depth of sample collection may also influence results (i.e. surface vs. bottom vs. water column) (Bellemain et al., 2016, Lacoursière-Roussel et al., 2016a). Water velocity could also influence detection by controlling residence time of water and associated eDNA accumulation rates (Pilliod et al., 2014). Water velocity is known to vary within a cross-section and among hydraulic units. Given the dynamic nature of rivers, it seems likely that the time or point location of sample collection will influence eDNA metabarcoding results. Little work has been done to test the influence of sampling location in rivers on results (Barnes and Turner, 2015). The current understanding of eDNA methodologies for rivers is primarily based on studies conducted in controlled experiments or small temperate rivers in the European Union and United States. Relatively little work has been done to test and adapt these methods for larger rivers in warmer climates (but see Robson et al. (2016) for tests of filters with large pore size for use in turbid tropical waters).

The current study tests the influence of sampling location and collection time within a river channel on eDNA metabarcoding results for a subtropical river. The inquiry presented here furthers the understanding of whether a sample collected at a single time and point location will be representative of the full suite of eDNA present at a site. This study seeks to expand the current knowledge of best practices for eDNA sample collection, by testing the molecular technique in a subtropical river. This work advances current knowledge by comparing results from replicate surface water samples collected at three different times of day, from several points along a cross-section and from two different hydraulic unit types. The findings can be used to improve future eDNA studies, by furthering our understanding of the influence of local conditions of rivers on

detection probabilities and metabarcoding results. The results can be used to improve field sampling strategies, and to inform methodologies that provide more representative samples from field collections. Primer design, amplification (PCR), and sequencing are known to introduce bias that may influence richness and beta diversity estimates from eDNA metabarcoding data (Smith and Peay, 2014). Two primer sets, that both amplify a broad range of taxa, were used in this study: the Teleo primers (Valentini et al., 2016) which are designed to amplify bony fish, and the universal eukaryote primers for the 18S mitochondrial region (Jarman et al., 2013). The work presented here tested three hypotheses: 1) the point location of sampling along the surface of a cross-section will influence richness or assemblage estimates, 2) the hydraulic unit type has an impact on richness or assemblage estimates, and 3) the time of sample collection has an impact on richness or assemblage estimates.

5.3 Methods

5.3.1 Field Sampling

Samples were collected in the Brisbane River, southeast Queensland, Australia (Figure 5.1a) over one day (29 February 2016). Cross-sections of two sequential hydraulic units, one pool and one riffle, were sampled (Figure 5.1b).

5.3.2 Site Survey

Cross-sections of the pool and riffle habitats were surveyed using a Sokkia CX5 Total Station set to an arbitrary datum of 100 meters. The true (when facing downstream) left edge of water (LEW) and right edge of water (REW) were surveyed and a tape was stretched across the channel so that subsequent velocity measurements could be spatially located relative to the surveyed cross section. Using a SonTek

Flowtracker Handheld ADV, velocity profiles were measured at stations spaced at no less than 1/20 the wetted width of the river.

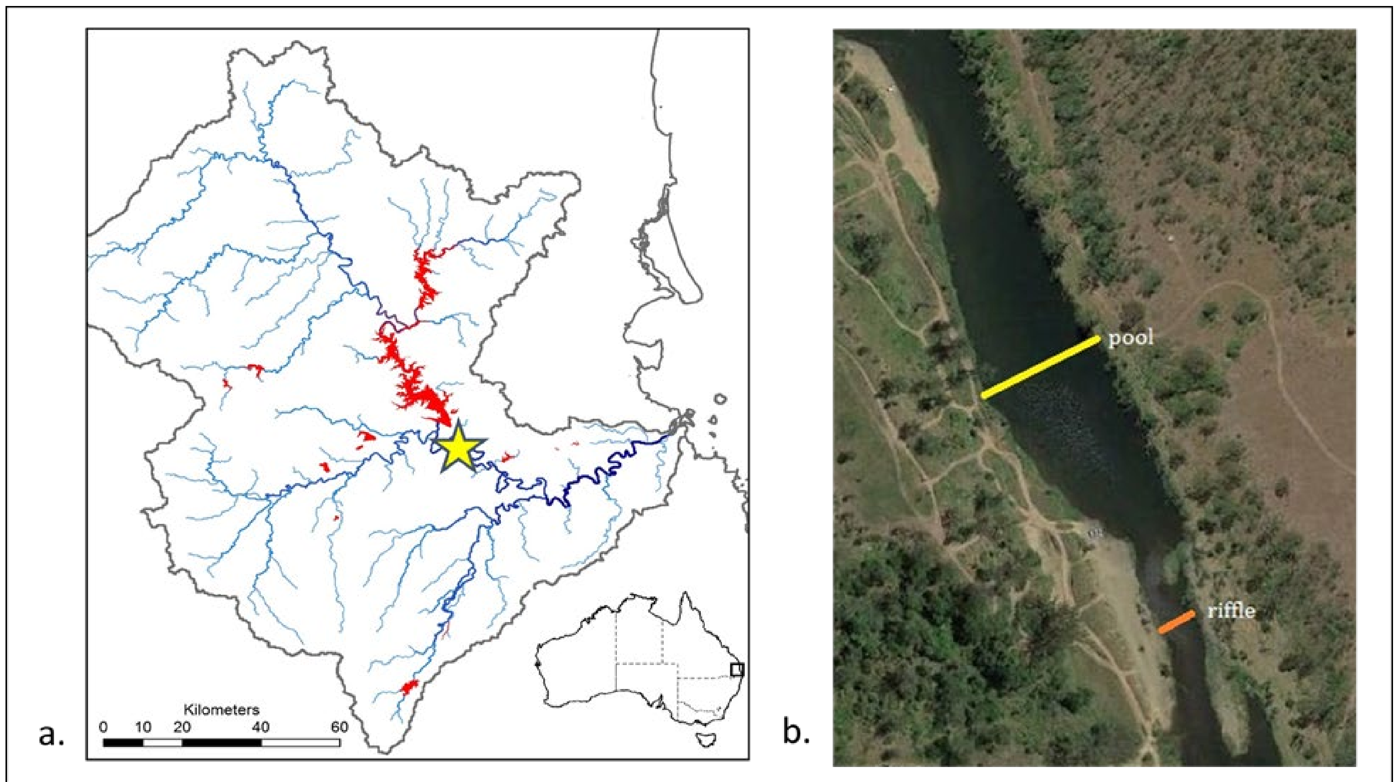

Figure 5.1: (a) Sampling location (star) within Brisbane River catchment, southeast Queensland, Australia; and (b) Cross-section locations for the pool (yellow line) and riffle (orange line) hydraulic sampling units

In areas of the cross-section where velocity was relatively faster than the surrounding area, measurement stations were added to ensure that no more than 20% of the flow was represented by that measurement cell. Velocity profile measurements in the riffle started at an initial depth of five centimetres below the water surface, and progressed in 10 cm increments until the bed of the river was reached. In the pool, velocity profile measurement started at an initial five centimetres and then progressed in 25 cm increments until the bed of the river was reached. The final bed velocity measurement was taken approximately five centimetres above the bed. Additional surface velocity measurements were recorded at 1200 hrs at each of the locations

where water was sampled for eDNA collection (Table 5.1). Discharge of each cross-section was calculated and compared against the reported discharge of the river to ensure that discharge differences were less than 10%. Isovel (velocity contour) maps of both the pool and riffle were constructed by creating isovel contours bounded by the cross-section in Matlab (Version R2017a) (Natick, 2017). Contours were interpolated using nearest neighbour methods, assuming that the bed of the river had a velocity equal to zero. Interpolated contours were then smoothed by fitting a two-dimensional spline to the contours.

5.3.3 eDNA Sample Collection

For each hydraulic unit, samples were collected from the surface at three points along the cross-section: (A) at the river bank, (B) at a distance of 25% of the wetted-channel width from the bank, and (C) at the center of the channel (a distance of 50% of the wetted-channel width from the bank) (Figure 5.2). A canoe was used to collect samples from points B and C in the pool. For each sampling point along a cross-section, samples were collected at three time points throughout the day (06:00, 12:00, and 18:00). Samples collected from some time points were not included in the analyses due to poor quality of sequencing results. Field technicians wore gloves at all times, and sampling equipment was sterilized with 1.5% Sodium Hypochlorite solution, including waders and canoe.

Three sterile 2-L plastic jars were flushed three times with river water before simultaneously being filled for collection from the surface at each sampling point. Each 2-L jar was shaken vigorously for 3 seconds, and the contents were rapidly poured over the top of a sterile 15-mL falcon tube until the 2-L jar was empty. Any excess water in the tube above the 15-mL measuring line was tipped off. 15-mL tubes that had been pre-filled with PCR-grade water inside a laminar flow hood were opened at the field site

and treated as river water samples throughout the molecular workflow to monitor contamination and serve as negative controls for extraction and PCR. The water in 15-mL tubes was transferred in the field to sterile 50-mL falcon tubes that had been pre-filled under a laminar flow hood in an ancient DNA facility with 33 mL of 100% EtOH and 1.5 mL of 3M sodium-acetate (Valiere and Taberlet, 2000). Tubes were kept at 0°C on ice in a portable freezer for a maximum of 14 hours until transport to the laboratory. They were stored at -80°C until extraction.

5.3.4 Laboratory Methods

5.3.4.1 eDNA Extraction

Extraction methods followed an alcohol precipitation protocol, as is used by many other eDNA studies, based on methods from Ficetola et al. (2008). 50-mL tubes were centrifuged at 3220rcf for 60 minutes at 6°C, and the supernatant was discarded, taking care not to lose the pellet. For sample R019A, about 15 mL leaked out in freezer before centrifuging, so 15 mL of 100% EtOH was added to this sample tube to balance it for centrifuging. Tubes contained a small amount of EtOH after this process which was removed by freeze-drying overnight. 100 µL of elution buffer (Qiagen, Valencia, CA) was added, tubes were vortexed on high for 5 seconds, centrifuged at 2000rcf for 5 seconds, and incubated at room temperature for 15 minutes. 80 µL of the solution was aliquoted into a sterile 2-mL tube, and the remaining solution was aliquoted into tubes for storage. 700 µL of SET buffer (50 mM Tris-Cl pH 8.0, 100 mM NaCl, 1 mM EDTA), 20 µL of Proteinase K (20mg/ml, Qiagen, Valencia, CA), and 70 µL of 10% SDS were added to each 2-mL tube and they were incubated on a rotor overnight at 56°C. 700 µL of phenol:chloroform:isoamyl alcohol was added to each tube, they were vortexed on high for 5 seconds, and spun at 20238 rcf for 5 minutes in a benchtop centrifuge. The top 400 µL was aliquoted into a fresh tube, 400 µL of chloroform was added, they were vortexed

at full speed for 5 seconds, then spun at 20238 rcf for 2 mins in a benchtop centrifuge. The top 300 µL was aliquoted to a fresh tube, 10 µL of 3M sodium acetate and 10 µL of 0.25% LPA were added, the tubes were vortexed for 2 seconds, and centrifuged at 2000 rcf for 10 seconds. 300 µL of isopropanol (pre-chilled to -80°C) was added, and the tubes were shaken by hand to mix. Tubes were chilled at -80°C for one hour, and spun at 20238 rcf for 15 mins. The supernatant was carefully tipped off. 500 µL of 80% EtOH was added to each tube and mixed by hand, and tubes were spun for 2 mins at 20238rcf. As much EtOH as possible was removed with sterile pipette tips, and tubes were left open in a laminar flow hood with the fan on overnight (9 hours) until all EtOH had evaporated. The pellet was re-suspended in 40 µL of elution buffer (EB buffer, Mat No. 1014608, Qiagen, Valencia, CA), and extract was stored at -20°C.

5.3.4.2 Standard PCR

The Teleo_F (5' – ACACCGCCCGTCACTCT - 3') and Teleo_R (5' – CTTCCGGTACACTTACCATG – 3') primer pair were used (Valentini et al., 2016) to amplify bony fishes. To amplify sequences for a wider range of taxa, a universal 18S primer for all eukaryotes (SSU39'F: 5' - CACCGCCCGTCGCTACTACCG – 3'; and SSU39'R: 5' – GGTTCACCTACGGAAACCTTGTTACG – 3') was also used (Jarman et al., 2013). All primers were fusion primers that contained Nextera indexing sequences at the 5' end, as described in the Illumina Nextera protocol (Illumina, 2016). All pre-PCR preparations were carried out in a laminar flow hood, and all equipment was regularly sterilized with bleach and UV-light. PCR conditions were as follows: each reaction contained a volume of 22µL, with final concentrations of 1x supplied buffer (New England Biolabs, Inc., Ipswich, MA, USA), 0.2mM dNTPs, 0.5mg/mL BSA (New England Biolabs, Inc., Ipswich, MA, USA), 0.05µM of each forward and reverse primer, and 0.1 units of Q5 hot start polymerase (10units/mL, New England BioLabs, Inc., Ipswich, MA, USA). Each reaction

contained 1μL of eDNA extract. Thermal cycling was at 98°C for 30 sec (x1 cycle); 98°C for 10 sec, 55°C (Teleo T_m) or 69°C (18S T_m) for 10 sec (x26); 72°C for 6 sec, final extension at 72°C for 30 sec, and a 10°C hold until removed and stored at -20°C. Both primer sets were used on every sample in separate reactions. Negative extraction and PCR controls were used to monitor potential contamination throughout the molecular workflow. For negative PCR controls PCR-grade water was added instead of eDNA extract. Two PCR replicates were performed for each sample and negative control, and products were pooled to be used as the template for a second round of PCR for annealing Illumina adapters.

5.3.4.3 Illumina Library Preparation and Sequencing

Nextera adapters were added by thermocycling under the following conditions: 98°C for 30 sec (x1 cycle); 98°C for 10 sec, 59° x 26 cycles; 72°C for 6 sec, final extension at 72°C for 30 sec, and a 10°C hold until removed and stored at -20°C. Products from positive reactions and negative controls for each primer set were pooled. Pools for each primer set were cleaned separately with Agencourt AMPure XP beads (Bekman Coulter, Inc. Indianapolis, IN) following the manufacturer's protocol, except that a ratio of bead to product volume of 1:1 was used for Teleo primers and a ratio of 0.9:1 was used for the 18S primers. The size of the clean products was determined on a Bioanalyzer DNA 1000 chip (Agilent Technologies, Santa Clara, CA, USA) and quantified with a Qubit (1.0) fluorometer (Agilent) and qPCR using the NEBNext Library Quant Kit for Illumina (New England Biolabs, Inc., Ipswich, MA, USA), following manufacturer's instructions. Based on qPCR and bioanalyzer quantitation results, an equimolar amount of the sample pool for each primer set was added to a final pool combining the product from the two primer sets. This final pool was diluted with molecular grade water to 4nM for paired-end sequencing on an Illumina MiSeq at the Griffith University DNA Sequencing Facility

following the manufacturer's protocol (Nextera XT DNA 150 kit, Illumina, Inc. San Diego, CA).

5.3.5 Bioinformatics Methods

Sequence reads were demultiplexed by sample using the MiSeq native bcl2fastq2 command. Reads were merged using PEAR v0.9.10 (Zhang et al., 2013), then quality filtered and dereplicated using USEARCH version 8.0.1623 (Edgar, 2013) and then checked for chimeras, denoised and clustered into zero-radius OTUs (ZOTUs) using USEARCH v10.0.240 (Edgar, 2016). Taxonomic assignment was performed using a naive Bayesian classifier (Wang et al., 2007) against a custom database. The latter database was constructed by *in silico* PCR (Bellemain et al., 2010, Ficetola et al., 2010) using the primers from the present study (Table 1) to amplify homologous DNA regions from the NCBI nt database (accessed March 2018) allowing up to three base pair mismatches per primer.

5.3.6 Statistical Analysis

Variation in taxon richness (total number of OTU's per sample) and assemblage structure (number of reads per OTU, expressed as a percentage of total reads for each sample) of the fish and eukaryote datasets were analyzed using Permutational Analysis of Variance (PERMANOVA) (Anderson, 2001) using Primer, ver. 6 (Clarke and Gorley, 2006), (see http://www.primer-e.com/) with the PERMANOVA add-on package (Anderson et al., 2008). A type III partitioning of sums of squares was used to accommodate the unbalanced nature of the data (Anderson et al., 2008), as samples from some time points were missing from particular habitat and transect point combinations. A factorial design with the following main effects was analyzed: Habitat Unit, Sampling Point and Time. Analysis of variation in taxon richness was based on a Euclidean distance matrix, which provides a univariate statistical test similar to

Analysis of Variance (ANOVA) (Anderson et al., 2008). Analysis of variation in assemblage structure was based on a Bray–Curtis similarity matrix for the PERMANOVA.

Non-metric multidimensional scaling ordination (Clarke and Warwick, 2001, Clarke, 1993) was used to summarize variation in assemblage structure. The taxa that correlated (Pearson's r. > 0.5) with sample position in ordination space were overlaid as vectors and used to interpret the ordinations.

For each dataset (eukaryotes and fish), we ran analyses on the subset of OTU's with a read count ≥ 10. Removal of OTU's with a low read count (<10) was undertaken to control for the effect of potential incorrect assignation of occasional sequences to the wrong sample (Schnell et al., 2015, Valentini et al., 2016). Prior to all analyses of fish data, non-fish OTUs (mammals, amphibians & reptiles) were removed.

5.4 Results

5.4.1 Velocity

Velocity was highly variable both within the pool and the riffle, and between the pool and the riffle (Figure 5.2). For the pool cross-section, surface velocity at the points where eDNA samples were collected varied from a minimum of 0.001 cms (Point A), to a maximum of 0.11 cms (Point B) (Table 5.1). For the riffle cross-section, surface velocity varied from a minimum of 0.012 cms (Point A) to a maximum of 0.756 cms (Point C) (Table 5.1). Therefore, among the locations where eDNA samples were collected, point velocity at the surface varied by up to a factor of 100 for the pool and 63 for the riffle (Table 5.1). Velocity also varied with depth, and contours are shown for the full cross-sections in Figure 5.2.

For the pool, the fastest flow recorded (0.3cms) can be seen as the contours between points A and B, starting just below the surface and dropping down into the

deepest part of the channel (Figure 5.2). This current may be carrying eDNA from a further distance upstream, while the water moving very slowly near the river's bank (0.01cms) may be aggregating eDNA from within the pool. For the riffle, the fastest flows were found closer to the center of the channel (0.87+cms), while the lowest flows were also near the bank (Figure 5.2)

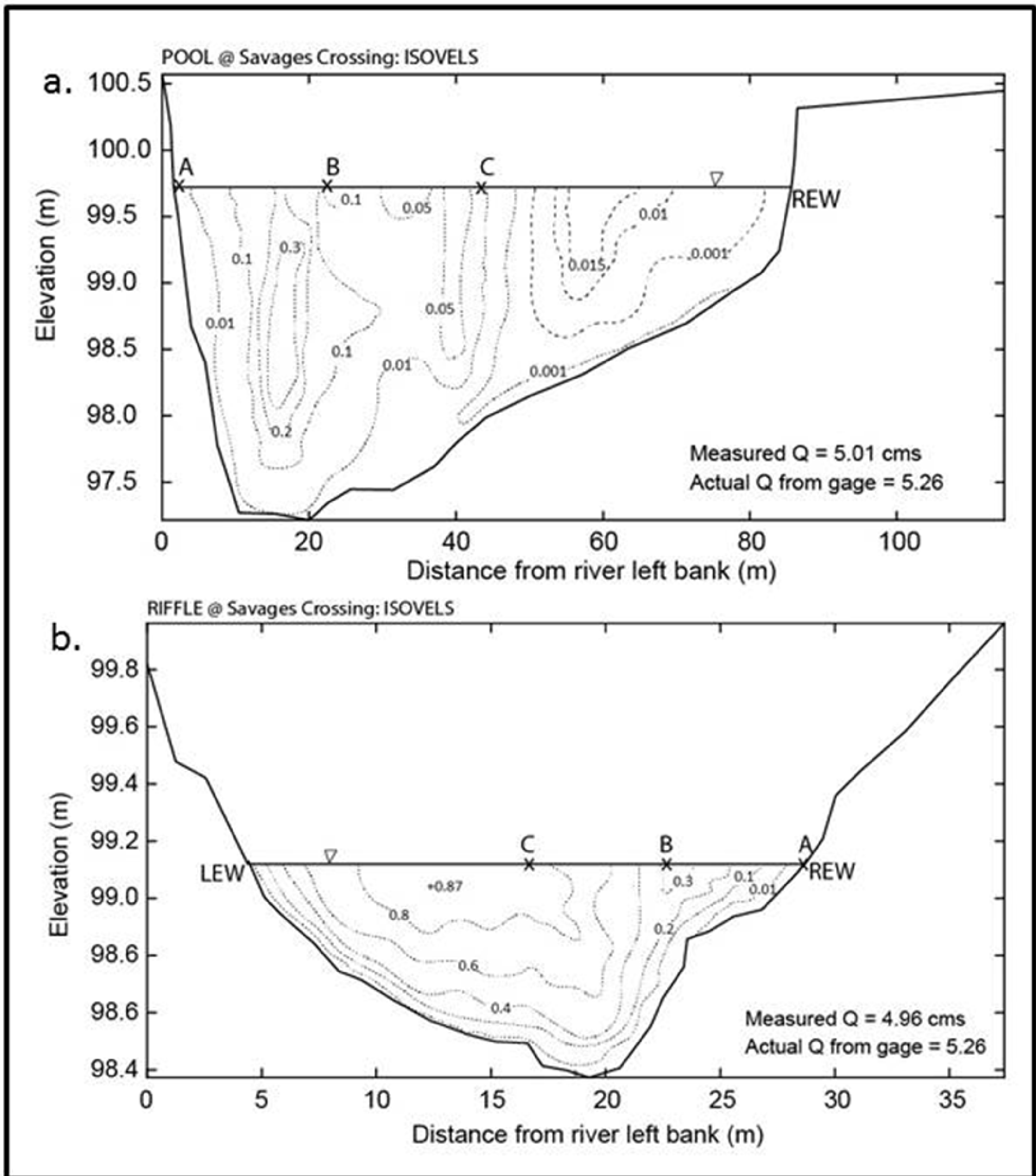

Figure 5.2: Isovel profiles (water velocity contours (cm/s) for the pool (a) and riffle (b) cross-sections, showing the point locations for eDNA sample collection (A, B, and C). The triangle indicates the water surface line.

Table 5.1: Velocity of surface water for a pool and downstream riffle at points on transects where eDNA samples were collected. Point A is at the river bank, Point B is 25% of the wetted channel width, and Point C is 50% of the wetted channel width along a straight transect from left bank to right bank. Numbers presented were collected at 1200 hours.

	Velocity (cms)		
	Point A	Point B	Point C
Pool	0.001	0.11	0.027
Riffle	0.012	0.197	0.756

5.4.2 eDNA metabarcoding

5.4.2.1 Fish

Amongst all samples, a maximum of 6 and minimum of 1 fish OTUs (mean = 2.4, SD = 1.6) were detected (Figure 5.3). The number of fish OTUs found in samples collected at different transect points, times, and hydraulic units showed substantial variation in richness among samples (Figure 5.3).

Fish assemblages based on the proportion of reads per OTU also varied substantially among samples (Figure 5.4). Some samples were dominated by a single order, and only one set of triplicate samples (noon, transect point C, pool) showed somewhat even distribution of proportion of reads among several taxa detected in those samples (Figure 5.4).

Notwithstanding the considerable variation in fish eDNA metabarcoding results among samples, there was no significant variation (P>0.05) explained by hydraulic unit, sample point or time (Table 5.2). Ordination plots showed taxonomic assemblages did not cluster by the factors tested in this study (Figure 5.7).

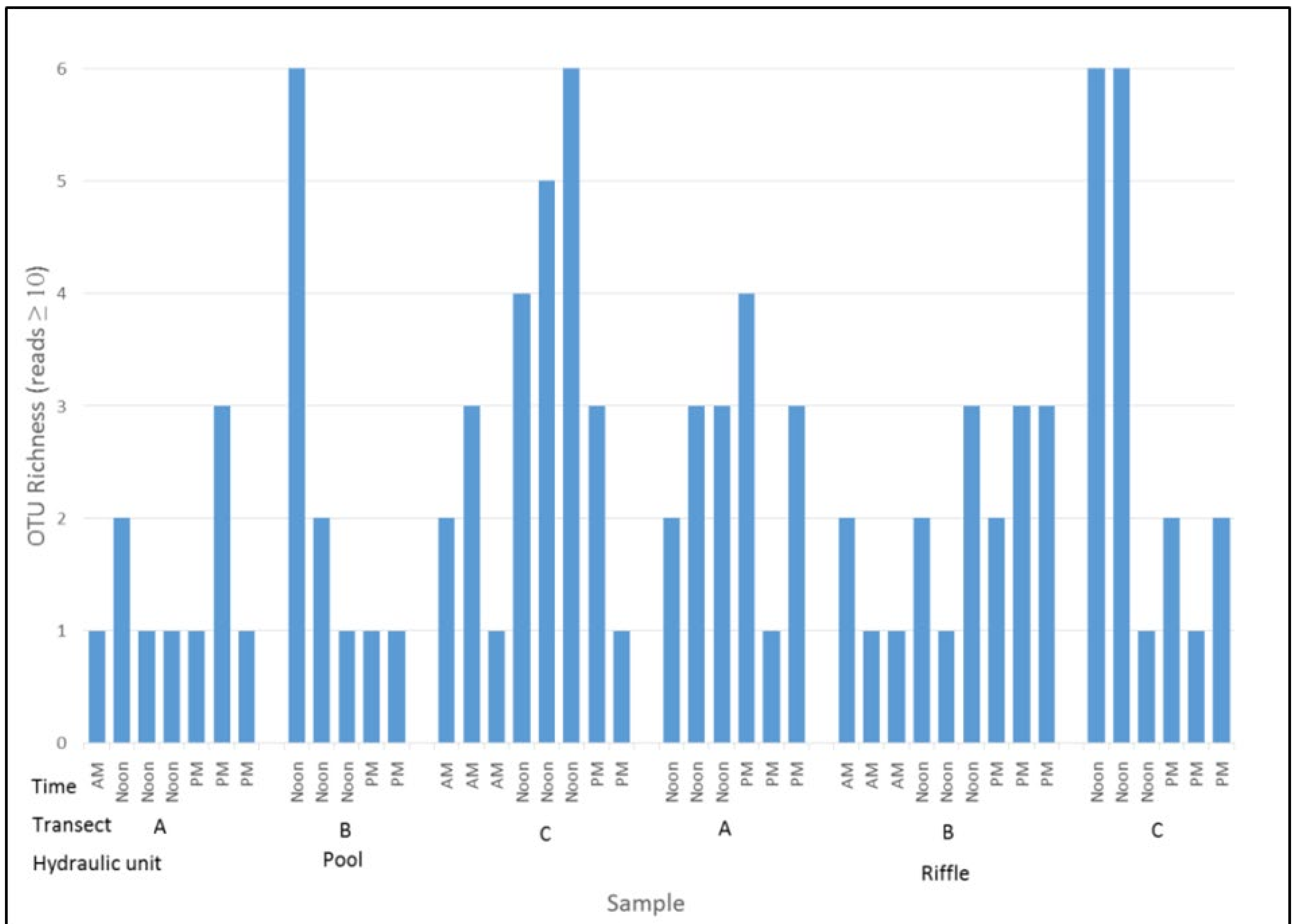

Figure 5.3: Number of fish OTUs for samples collected at different times, transect points, and hydraulic units, when only OTUs with a read count ≥10 are included, based on metabarcoding with generic fish primers

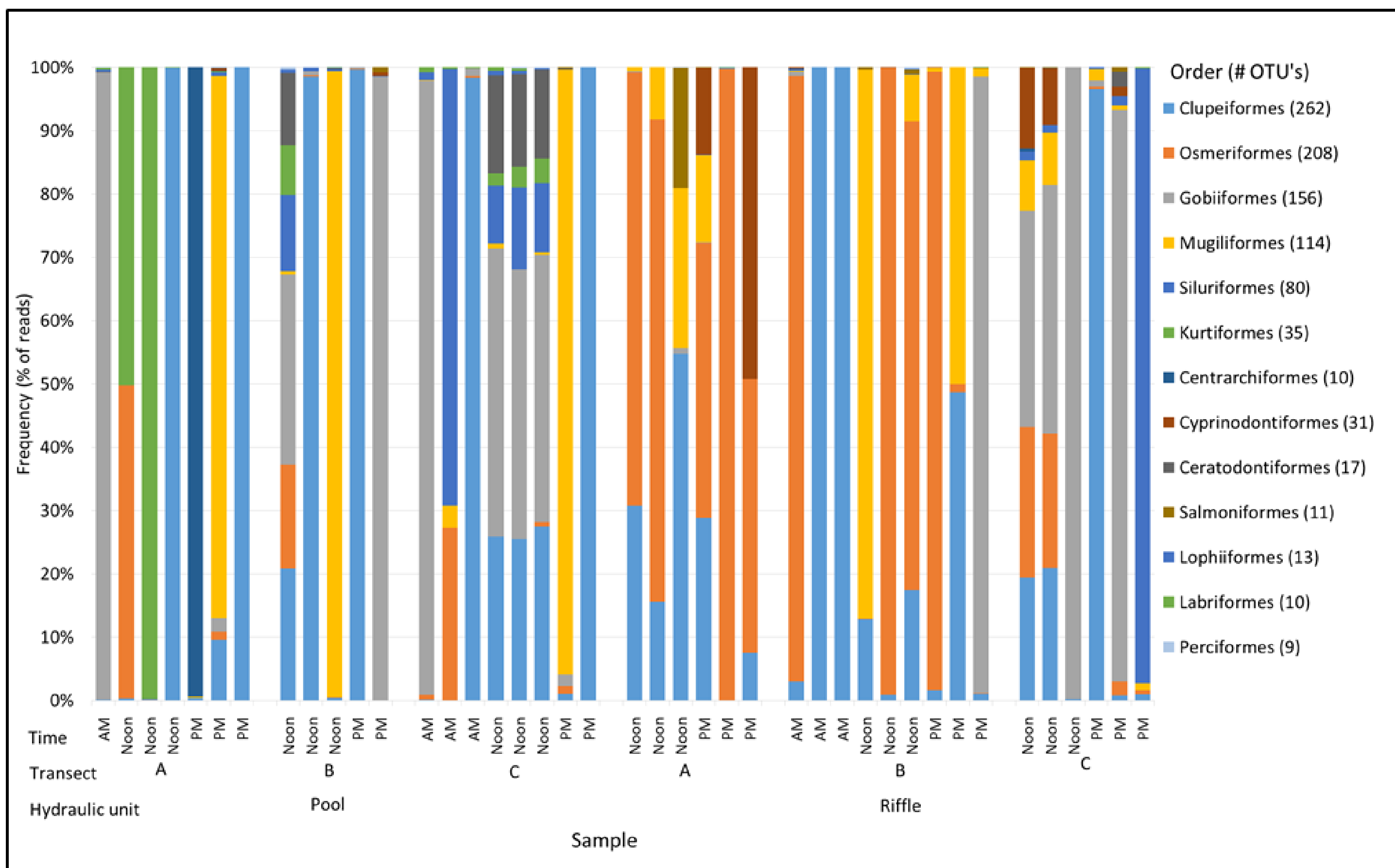

Figure 5.4: Variation in fish assemblages from samples collected at specific times, transect points, and hydraulic units. Data are based on proportion of reads assigned to OTUs within each taxonomic Order of fishes. The total number of OTUs within each Order is shown in parentheses.

Table 5.2: P values from nested permutational multivariate analysis of variance examining variation in fish and eukaryota taxon richness and assemblage structure, respectively, in response to factors Hydraulic Unit (HU), Channel Position (CP) and Sampling Time (T). P-values obtained with 999 permutations; those < 0.05 are underlined.

	df	Taxon Richness	Assemblage structure
Fish			
HU	1	0.074	0.091
CP	2	0.066	0.205
T	2	0.120	0.261
HUxCH	2	0.274	0.497
HUxT	2	0.291	0.217
CHxT	4	0.210	0.131
HUxCHxT	2	0.348	0.110
Residual	26		
Total	41		
Eukaryote			
HU	1	0.455	0.008
CP	2	0.432	0.075
T	2	0.051	0.002
HUxCH	2	0.539	0.001
HUxT	2	0.437	0.002
CHxT	4	0.426	0.001
HUxCHxT	2	0.198	0.049
Residual	31		
Total	46		

5.4.2.2 Eukaryotes

Amongst all samples, a maximum of 27 OTUs and a minimum of 12 OTUs (mean = 20, SD = 3.7) were detected (Figure 5.5). Phyla assemblage based on proportion of reads showed substantial variation among samples (Figure 5.6). There were four phyla that were consistently dominant in read proportions in all samples (Figure 5.6).

PERMANOVA results examining variation in taxon richness and assemblage showed that hydraulic unit type, channel position, and sampling time had significant effects on

assemblage structure, but not taxon richness (Table 5.2). The significant 3-way interaction term indicated that assemblage structure differed substantially between samples from each combination of factors. Ordination plots of variance in assemblage structure among samples indicate some grouping of samples by hydraulic unit type, and there appears to be some distinction between results for the pool and the riffle (Figure 5.7).

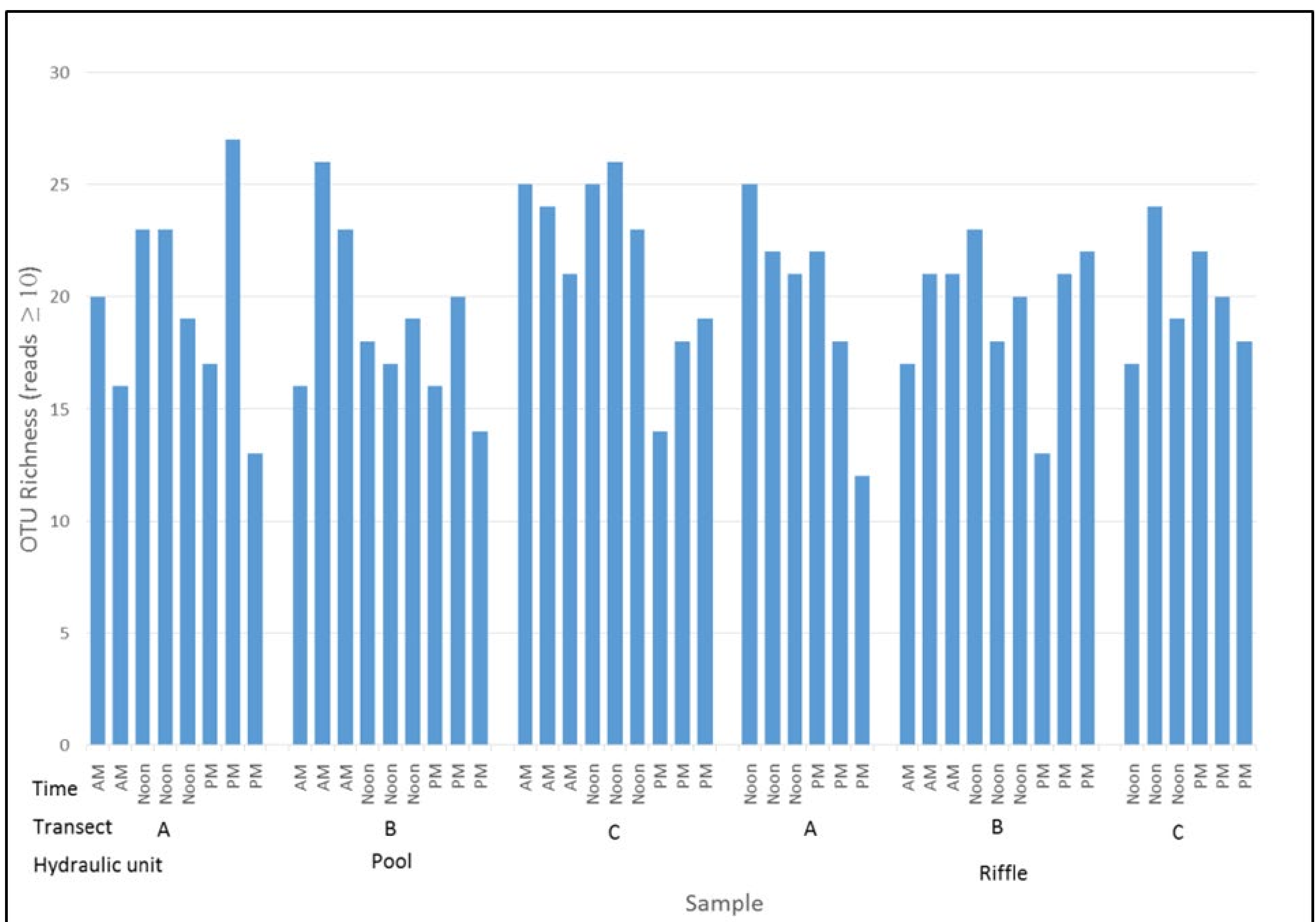

Figure 5.5: Number of eukaryote OTUs for samples collected at different times, transect points, and hydraulic units, (a) for all OTUs and (b) when only OTUs with a read count ≥10 are included, based on metabarcoding with generic eukaryote primers.

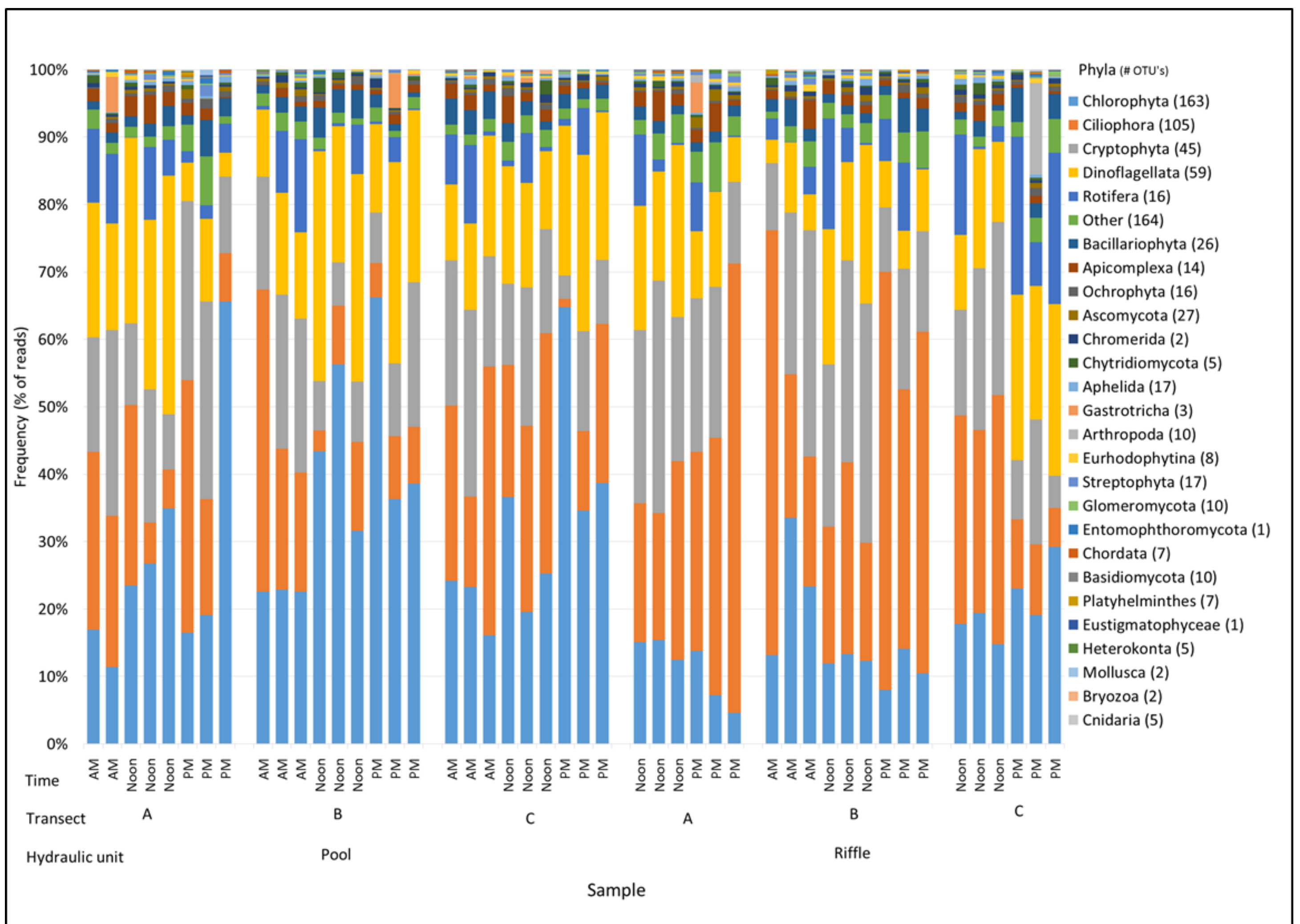

Figure 5.6: Assemblage of Eukaryotes from samples collected within specific times, transect points, and hydraulic units based on proportion of reads assigned to different phyla.

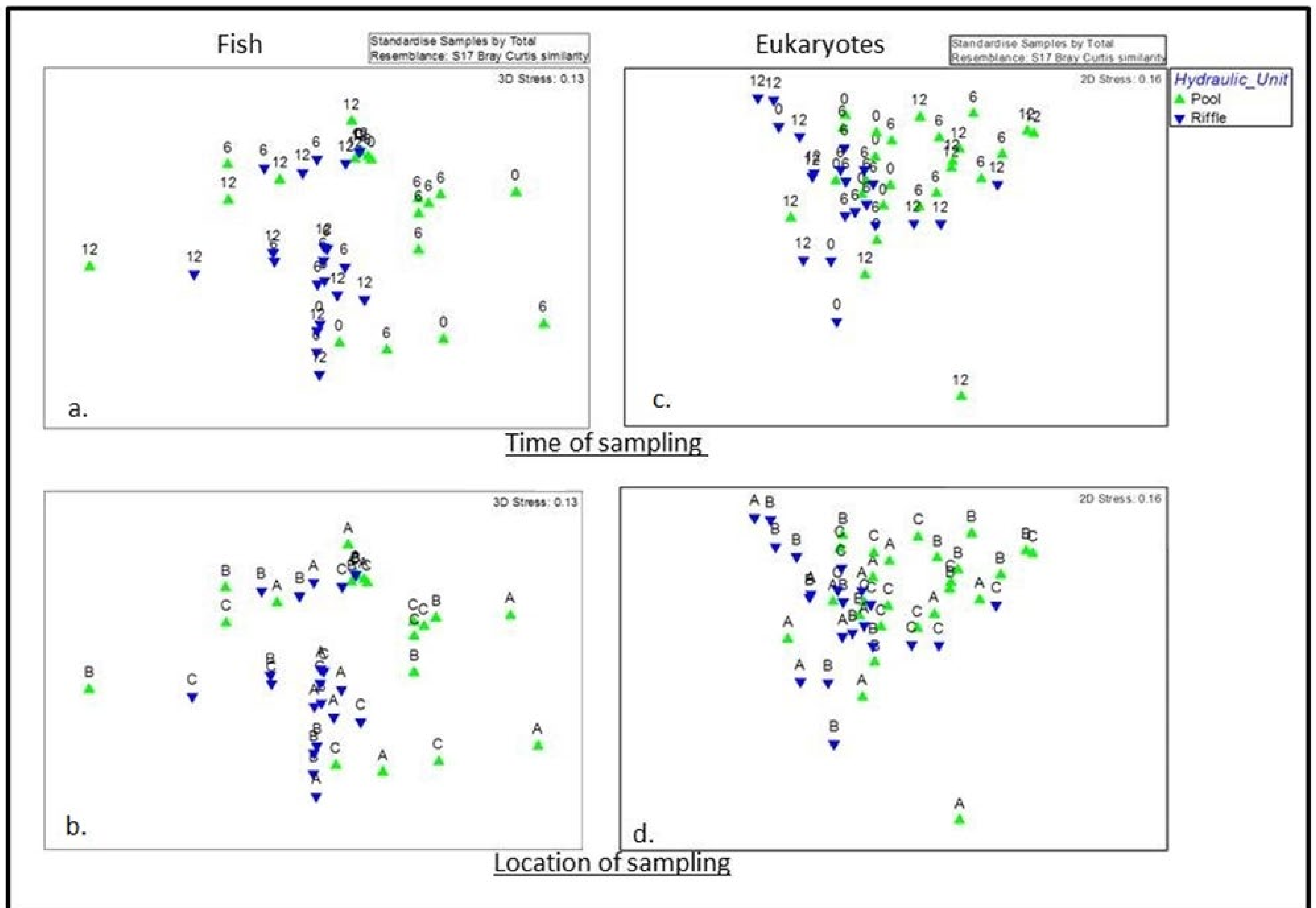

Figure 5.7: Ordination plots for fish (a. and b.) and for eukaryotes (c. and d.), with labels showing the time of sampling (a. and c.) and channel position location of sampling along transects (b. and d.). Green triangles represent data points for the pool hydraulic unit and blue triangles represent data points for the riffle.

5.5 Discussion

Based on a review of the literature (Chapter 2), this is the first time that an eDNA metabarcoding study has evaluated the influence of time, location of sampling point along a cross-section, or hydraulic unit type on the results of eDNA metabarcoding of water samples from a natural river. There are a few published studies that have considered the influence of velocity and substrate type on eDNA detection rates (Jane et al., 2015, Jerde

et al., 2016, Shogren et al., 2017) and at least two studies where samples were collected at different depths (Moyer et al., 2014, Bellemain et al., 2016). However, these studies used species-specific primers.

Water samples for eDNA studies are commonly collected from the river bank and it has been implicitly assumed in many studies that this provides representative eDNA samples for the site (Laramie et al., 2015). However, it has been shown that variables such as shading from bank vegetation and associated limits on UV light penetration and temperature influence eDNA degradation rates (Pilliod et al., 2014). Shading from bank vegetation is likely to cause eDNA degradation from UV light penetration to be lower near a vegetated bank compared to an unshaded area, but the locations and levels of shading will vary throughout the day. In the current study, the divergent eDNA metabarcoding results found among three sampling points (channel positions) support the conclusion that the choice of sampling location laterally across the river channel may influence eDNA metabarcoding results (Table 5.2). Some of the replicate samples from a single sampling point showed similar taxonomic assemblages, while others showed high variation (Figures 5.4, 5.6). This could partially be caused by the small sample size used, as only three replicate samples were collected for each time point. Surface velocity varies laterally across a river channel, and higher velocities can effectively act to dilute the eDNA of species that are present at a sampling site by adding water and exogenous eDNA from upstream.

These findings indicate it may be necessary to collect more than three replicate samples per site in order to collect eDNA that is representative of the full assemblage of fauna present. If the point of sample collection along a cross-section had no influence on results, then we would expect to see little variation in the number of OTUs detected from samples collected at points A, B, or C. However, the results showed there was substantial

variation in OTU count among samples collected at different points across the channel. Yet for the results identified with the fish primers, none of the factors considered in the model had significant effects. Some alternate factor must be causing the variation in richness and assemblage that is apparent in Figure 5.3. Alternate factors may include water temperature, depth, velocity, UV light penetration, variation in eDNA shedding rates among taxa, and behavior of organisms, some of which vary laterally with distance from the river bank. It is possible that organisms with higher densities are more likely to be uniformly detected across samples, while stochastic results could indicate that an organism is present at low density. For the results identified with the eukaryote primers, the model showed three factors had significant effects on assemblage structure (Table 5.2). Hydraulic unit type, channel position within the hydraulic unit, and time of collection for a certain channel position within a hydraulic unit all had significant effects on taxa assemblage.

Overall, the results reported here suggest that richness was not significantly influenced by the factors tested in the model, while results from eukaryote primers indicated that several significant effects of hydraulic unit, channel position, and time of sample collection influenced taxonomic assemblage structure (Table 5.2). In rivers, assessments of richness based on eDNA metabarcoding are likely more reliable than taxonomic assemblage estimates. Accurately assessing taxon assemblages with a single universal primer set remains challenging, and is more likely influenced by the highly variable physical environment of a natural river. The difficulty in determining taxa assemblages is also related to the additional bioinformatic steps required, compared to the steps needed to count richness. Assessing taxonomic assemblage requires assignment of taxonomic nomenclature to OTUs produced from sequencing, while richness can be calculated by counting OTUs. The identification of assemblages based on genetic

distances between OTUs does not require taxonomic assignment, and is an alternative to taxonomic assessment of assemblages. The bioinformatic steps required for taxonomic assignment require the availability of reference genomes with which to match OTUs identified through eDNA metabarcoding.

Many eDNA studies implicitly assume that the selection of hydraulic unit type for sampling does not have a substantial influence on results. Results from ordination plots for this study indicated some grouping by hydraulic unit type (Figure 5.7). These differences are likely attributable to the variation in velocity and depth between the pool and the riffle. For the pool, the highest flows (0.3cms) were between points A and B (Figure 5.2). This current may be carrying eDNA from further upstream, while the water moving very slowly near the river's bank (0.01cms) may be aggregating eDNA from within the pool. For the riffle, isovel profiles showed a fast current near the center of the channel. This current is likely carrying eDNA from the pool upstream and/or mixing the eDNA present throughout the riffle. Some literature indicates that results are more reliable from small to medium enclosed water bodies if sample collections are spaced representatively throughout the sampling area and sample volume is sufficient (Takahara et al., 2013, Kelly et al., 2014, Robson et al., 2016, Bista et al., 2017). One study in a small reservoir detected a variance in richness and assemblage estimates from samples collected near an inflow carrying stream water, while results throughout the rest of the reservoir were more homogenous (Evans et al., 2017a).

It is likely that taxa composition is truly different between the pool and riffle, as some taxa are likely to be associated with one habitat type over the other (Kennard et al., 2006). However, given that the riffle was immediately downstream of the pool and that eDNA is detectable for a minimum of several hours after shedding (Dejean et al., 2011), some of the eDNA shed in the pool could be detected in the riffle through downstream

transport. The differences in assemblages among hydraulic units mean that certain taxa will be a more predictable component of the metabarcoding assemblage in this analysis factor. Alternately, organisms that are uncommon in riffles should be a more stochastic component of the eDNA assemblage in a riffle rather than other hydraulic unit types. The results indicate that comprehensive eDNA sampling should include several types of hydraulic units and that information about what type of habitat was selected for sampling should be reported in publications.

Techniques for representative water quality sampling and knowledge of particle transport dynamics in rivers have been developed (Eaton et al., 1998), and recent work can provide useful insight into how the choice of sampling location within the river channel may influence eDNA detection (Shogren et al., 2017). Broader consideration of the highly dynamic nature of rivers and the potential influence of velocity on detection probabilities is warranted. The magnitude of variation in velocity is different among hydraulic unit types, so the choice of unit type for sampling should be explicitly considered and tested in study design and reported together with other results. Isovels are more variable in larger rivers; therefore, the level of influence velocity exerts on detection probabilities is likely to vary with river class.

The results of this study indicate that the highly dynamic nature of currents and velocities in different hydraulic units of a river, and along a single cross-section within a unit, may influence eDNA detection probabilities and eDNA metabarcoding results. The strength of this influence likely varies according to river class and discharge levels, and it is expected that the power of this bias would be lower in small streams. This study indicates that assemblage estimates based on eDNA metabarcoding are influenced by the location of sampling within the river channel (both along a cross-section and by hydraulic unit type). To increase the taxonomic comprehensiveness of eDNA metabarcoding

surveys for assessing richness or taxonomic assemblage in rivers, a stratified and integrated sampling approach similar to that used for water quality assessments for viruses and bacteria is recommended. Methods should typically involve collecting samples either from the center of the channel or the point where velocity is highest, collecting samples at several different depths and points along the surface of a cross-section, and collecting samples from locations where deposition tends to occur, as stratified samples that are either aggregated or analyzed separately depending on research goals (Eaton et al., 1998). Recommendations made in an eDNA sampling protocol for streams published by the U.S. Forest Service advise collecting samples from the center of the river channel rather than the bank (Carim et al., 2015).

Recent progress in technologies associated with non-invasive genetics, such as eDNA metabarcoding, have provided improved conservation understanding (Beja-Pereira et al., 2009). Endangered and threatened species are notoriously difficult to sample through observation-based techniques, primarily because of their rarity. It is arguable that the highest conservation potential for eDNA methods lies in applications to large tropical rivers. eDNA studies in large tropical rivers are limited, relative to the large number of studies conducted on smaller temperate rivers. Additional work is needed to understand how detection probabilities vary in large tropical systems. Such rivers have the highest species diversity, typically have limited data available on richness and assemblages, and are especially difficult to sample with conventional methods. These systems are subject to increasing development pressures, with associated habitat degradation and high extirpation levels (Dudgeon, 2011). A recent review by a global group of eDNA experts concluded that this method is most advantageous when conventional methods are logistically difficult, have low detection probabilities, or are very costly (Goldberg et al., 2016). This describes the situation for many large tropical

rivers, and is a primary cause for the current lack of knowledge on the biodiversity of these systems. However, even if a method is useful from a time/cost perspective, its application is only as useful as its fidelity. Assessing sample-to-sample variance within a site and factor level, as was done in this study, advances our understanding of the fidelity of eDNA metabarcoding techniques. With further work to refine the method to better account for the influence of isovels, tropical conditions, and high discharge levels, eDNA metabarcoding can be a useful tool for recording and conserving aquatic species richness and assemblage in biodiversity hotspots across the planet.

CHAPTER 6: eDNA METABARCODING OF SEDIMENT CORES COLLECTED IN MORETON BAY, QUEENSLAND

6.1 Summary

Understanding past and current distributions of aquatic species is essential for conservation, yet notoriously challenging. Ancient environmental DNA (eDNA) preserved in sediments accumulated at the outflow of a river may enable the construction of a chronology of species' occurrences within a catchment. By correlating changes in a biotic assemblage with any changes in environmental variables such as shifts in land use, we can better understand how environmental changes may influence the potential for different species to occupy a river habitat and persist within it. In Eastern Australian rivers, increased sedimentation rates (Coates-Marnane et al., 2016) resulted from the rapid land-use change that occurred with settlement of Europeans (Capelin et al., 1998).

This chapter describes an exploratory effort to determine whether changes in land use are associated with shifts in taxonomic assemblage of fishes, by attempting to identify fish eDNA from sediment cores collected at the outflow of the Brisbane River, Queensland, Australia. eDNA was extracted from layers of sediment that were dated to pre-European arrival, post-European arrival, or modern periods. Metabarcoding was conducted using universal fish primers, yet the majority of sequences identified were assigned to bacterial reference sequences with low levels of support. The failure to identify fish eDNA may be the result of the relatively high proportions of bacterial DNA in samples compared to fish eDNA. Despite attempts to target and amplify only fish eDNA, the relatively large amounts of bacterial DNA in samples could cause PCR to be inefficient and result in PCR product that contains non-target sequences. Results may also indicate that fish eDNA is not preserved well enough to enable detection. High DNA degradation rates are expected to occur in the region given the warm climate (Hofreiter

et al., 2015). The chapter closes with hypotheses for why the method did not work, such as high eDNA degradation rates from the warm climate and high proportions of bacterial DNA in samples relative to fish eDNA. Conclusions also include recommendations for alternate laboratory methods that may enable successful identification of fish eDNA from subtropical sediments.

6.2 Introduction

All living organisms are continuously depositing DNA in their surrounding environment through the shedding of skin, mucous, hair, urine, feces, gametes, and other cells. Once this DNA is shed it is termed "environmental DNA" (eDNA) (Taberlet et al., 2012a). "Ancient DNA" typically refers to DNA extracted from sources such as sub-fossil bone, but can also be used to refer to cellular and extracellular DNA collected from ancient environmental samples such as sediment cores. Because eDNA is extracted from bulk environmental samples such as water or sediment, and the extractions will contain DNA from a wide variety of organisms such as plants, bacteria, and fungi in addition to the DNA of target species.

Sampling for eDNA involves collection of bulk environmental samples, for example water or soil, which contain cellular DNA, extracellular DNA, dissolved DNA, or fragments of DNA molecules that have been shed by an organism (Barnes and Turner, 2015). Previous studies have shown that after being shed, eDNA persists in the water column for approximately one to two weeks in ponds depending on environmental conditions (Dejean et al., 2011), and is no longer detectable in stream water one hour after an organism is removed from a site (Pilliod et al., 2014). However, much of this shed DNA is deposited on the substrate or adsorbed to sediment particles (Turner et al., 2015) and can be preserved for millennia under certain conditions (Haile et al., 2007, Thomsen and Willerslev, 2015). The amplification of ancient DNA from animal remains

was initially pioneered by Higuchi et al. (1984), who successfully extracted DNA from dried tissue of a museum specimen of the extinct Quagga (*Equus*). Research on remains found in ice, permafrost, and sediment followed (Hagelberg et al., 1994, Willerslev et al., 1999, Lambert et al., 2002, Willerslev et al., 2003, Willerslev et al., 2004, Ritchie et al., 2004)

The DNA preserved in sediment cores can serve as a chronology of shifts in biological community composition over time (Haile et al., 2007, Pedersen et al., 2016). Work to improve estimations of the timing of the extinction of megafauna in North America relied on this phenomenon (Haile et al., 2009). Work to understand prehistoric climate regimes and vegetation patterns also relies on this technique, typically examining shifts in the composition of flora types that are known to have variable responses or tolerances to changes in climate (Pedersen et al., 2013).

While humans have been present in Australia for approximately 60,000 years (Thorne et al., 1999, Bowler et al., 2003, Dennell and Petraglia, 2012), it is well understood that European settlement in the 1840s was followed by substantial changes in land use and associated impacts on ecosystems (Capelin et al., 1998). Extensive clearing of forests for pasture and farming led to large increases in sediment delivery into river catchments and receiving coastal areas, as well as increased nutrient inputs. Higher sedimentation rates and higher nutrient levels may play a role in influencing aquatic biodiversity in Eastern Australian rivers. The eDNA that is associated with sediments may serve as an archive of biological community compositions.

Most published eDNA research focuses on modern eDNA (Rees et al., 2014). There are several ancient DNA studies that use sediment cores (Willerslev et al., 2003, Haile et al., 2007) to track changes in biological communities (Rawlence et al., 2014).

For studies examining ancient vertebrate eDNA, nearly all are conducted in temperate or cold environments. The work presented here represents one of only a small number of attempts to amplify and sequence ancient environmental DNA from sediment cores collected in a subtropical environment. Other research has found substantially variable taxonomic composition along core profiles, which proves genetic material can be preserved in certain environmental conditions with stratigraphic integrity and over substantial time periods (Willerslev et al., 2003)

This chapter attempts to build on the above findings to evaluate whether DNA shed by freshwater organisms in a subtropical river catchment can be detected in marine sediment cores collected in an enclosed bay at the outflow of the river. It also aimed to identify whether land use changes upstream in the catchment (forest clearing for livestock/farming) were correlated with shifts in both freshwater and marine community structure. Generic fish primers were used in an attempt to amplify ancient eDNA from different layers of sediment dated from recent surface layers back to ~3500 years before present.

6.2.1 Study area

Moreton Bay is a shallow estuarine embayment located in the sub-tropical east coast of Australia (Figure 6.1). The Bay supports numerous habitats including seagrass meadows, mangrove forests, and coral reefs (Dennison and Abal, 1999). The bay is shallow, with an average depth of 6.8 m. It is bordered in the east by two large sand barrier islands, Moreton and Stradbroke, which restrict the tidal exchange of oceanic waters. Five major rivers drain into the Bay; these comprise a total catchment area of 21,220 km^2. The largest is the Brisbane River, which drains grazing and cropping lands in the upper catchment before passing through metropolitan Brisbane.

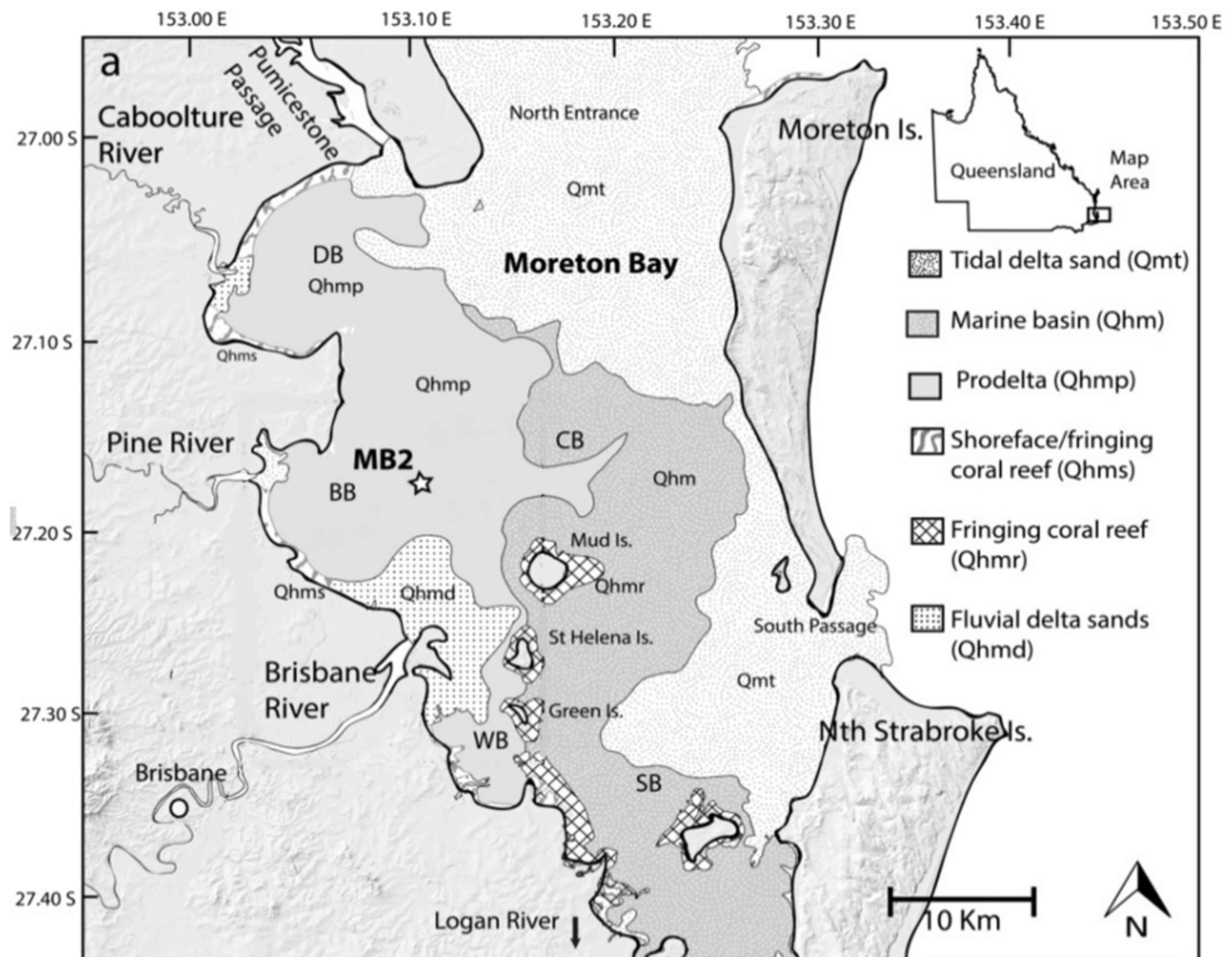

Figure 6.1: Map of Moreton Bay, Queensland, Australia with the location where cores were collected indicated by star, and the core sample location ID "MB2". Modified map from Coates-Marnane et al. (2016).

Following European settlement in the 1840s, the watershed draining into Moreton Bay experienced widespread vegetation clearing (Capelin et al., 1998). Only ~25% of the original native vegetation remains (Powell, 1998). Catchment erosion models have been used to describe the rate of increase in sediment input into Moreton Bay following catchment clearing, and typically suggest a 3 to 4 fold increase (Neil et al., 2002). The depositional basin of Moreton Bay is comprised of a mixture of fluvial and marine sediments (Maxwell, 1970). Fluvial delta sands and muds dominate western regions of the Bay, while clean marine sands and carbonate sediments dominate eastern, north eastern and south eastern regions. The area of clean sand facies within

Moreton Bay has been significantly reduced (by ~ 20%) since initial surveys in 1970 (Maxwell, 1970, O'Brien et al., 2012). This reduction is the result of an increase in the total proportion of fine sediments received by the Bay following European settlement, and is of critical concern to the health of tidal and sub-tidal ecosystems within the Bay (Coates-Marnane et al., 2016). The replacement of sands with fine clays coupled with a gradually shallowing basin is suggested to be gradually increasing the turbidity in western regions of the Bay (Coates-Marnane et al., 2016).

6.3 Methods

6.3.1 Sample collection

In November 2011, two sediment cores were collected from within the mud-dominated region of Moreton Bay directly north-north east of the mouth of the Brisbane River (Figure 6.1). A purpose built, barge-mounted hydraulic vibro-corer designed to sample sub-tidal sediment profiles was used to extract a long sediment core, referred to from here as MB2L. In addition, a short core was extracted from the less well-consolidated surficial sediments referred to from here as MB2S. This was done to ensure the recovery of the complete surface sediment profile that may not have been achieved using the vibro-corer alone. Divers using SCUBA, hammered a 15 cm diameter plastic pipe into the sediment and used suction to retrieve a section of sediment. Core sections collected were then stored in the dark at 4°C until sectioning for subsequent analysis.

Before sectioning, the entire outer layer (approx. 2mm thick) was removed with a sterile scalpel to eliminate material that may have been exposed to water or sediment from surface layers during core collection. The short core ("MB2S") was sectioned at 2 - 4 cm intervals and the long core (MB2L) was sectioned at 5 - 10 cm intervals and stored in individual vials at -20°C. Prior to extraction the vials were opened in a laminar flow

hood and parafilm was placed over the top with three small holes punctured using sterile syringe needles. The samples were freeze-dried and transferred to new sterile vials in a laminar flow hood using sterile pipette tips. All pre-PCR work was conducted in a lab that meets contamination control standards for ancient DNA.

6.3.2 Core stratigraphy

The two cores, MB2S and MB2L, consisted of organic rich muds with no visible stratigraphy and occasional shell fragments (Table 6.1). The exception to this was for the surface layer of MB2S (0-10cm) which comprised a distinct viscous clay-rich layer likely deposited following the 2011 flood event (Coates-Marnane et al., 2016).

6.3.3 Dating of the sediment cores

The full details of the dating of the sediment core are reported in Coates-Marnane et al. (2016). Ages of deposited sediments at site MB2 are summarized in Table 6.1. The presence of the fallout radionuclide ^{137}Cs in the sediment cores was used to date sediment to post-1959. Caesium-137 is a product of atmospheric nuclear testing that occurred from the 1950s to 1970s. In Australia, concentrations of fallout ^{137}Cs reached levels that are still detectable in soils and sediments from 1959, and the first appearance of ^{137}Cs in sediment cores is used to date sediment deposition to this time (Leslie and Hancock, 2008, Hughes et al., 2009). Pre-European sediments were identified using optical dating, which provides an estimate of when a grain of quartz sand was last exposed to the sun, and so determines an effective burial age (Aitken, 1998). Optical dating has been successfully applied in coastal sedimentary environments (Olley et al., 2004).

6.3.4 eDNA extraction

All steps for eDNA extraction were performed under a laminar flow hood in an ancient DNA facility. Approximately 0.25g dry weight of freeze-dried sediment was added to each bead tube from the MoBio PowerSoil extraction kit using sterile pipette

Table 6.1: Sample depth (cm), core stratigraphy, and burial ages of a selection of deposited sediments at site MB2. Sample depths shaded in purple were selected for sequencing of ancient eDNA.

Sample depth (cm) (short core "MB2S")	Sample depth (cm) (long core "MB2L")	Stratigraphy	Burial age
0-5			2011 Flood
5-10			2011 Flood
10-15			**Pre 2005**
	10-20	Uniform anoxic consolidated mud	Post 1950 (Cs137)
20-25			
25-30	20-30		
30-35			
35-40	30-40		
40-45	40-50		
55-60	50-60		Pre 1950
	60-70		
	70-80		
	80-90		
	90-100		
	100-110		480±40 (YBP)
	110-120		
	120-130		
	130-140		
	140-150		
	150-160		900±110 (YBP)
	160-170		
	170-180		
	180-190		
	200-210		1660±180 (YBP)

tips. The extraction followed the manufacturer's protocol except that the ratio of supernatant to binding buffer was altered by a factor of three, to 200μL of supernatant with 1200μL of binding buffer (C4 solution), in order to increase binding of short DNA fragments. This ratio was selected after a trial testing three different ratios of binding buffer (Appendix 6.1). Also, the volume of elution buffer was reduced to 50μL to obtain a higher concentration of DNA in final the elution. 25μL of elution buffer (C6 solution) was added to the center of each column filter, the columns were incubated at room temperature for 30 seconds and then centrifuged at 10,000g for 30 seconds, and then this process was repeated.

6.3.5 Primer selection and design

In vitro testing of universal fish primers (Table 6.2) from Valentini et al. (2016) (named "Teleo") and designed for this study (named "12SiiSHRT_R") were conducted with extractions of pure fish DNA from tissue samples. For *in silico* tests of primers, Geneious software alignment tools (version 8.1.1; Biomatters, Ltd) and a nucleotide BLAST of NCBI's GenBank (Benson et al., 2012) were used. All tests showed that fish extract amplified well at the expected fragment length and that primers were specific to fish.

Table 6.2: Primer sequences, melting temperatures for PCR

Primer Name	Primer Sequence	T_m (°C)	Reference
Teleo_F	ACACCGCCCGTCACTCT	56	(Valentini et al., 2016)
Teleo_R	CTTCCGGTACACTTACCATG		
12SiiSHRT_R	TTACCATGTTACGACTTGCCTCC	70	This study

6.3.6 Amplification and sequencing

Three rounds of PCR were performed in a nested PCR design (Gundersen and Lee, 1996), with round 1 amplifying fragments of fish DNA using the Teleo_F and

Teleo_R generic fish primers. The PCR conditions are described in Appendix 6.2. The product of the first round of PCR was used as the template for the second round. The second round of PCR used the Teleo_F primer and another fish-specific primer (12SiiSHRT_R) that targets a region of DNA 3' of the Teleo_R primer. These nested primers also had adapters to enable a third round of PCR to facilitate indexing for Illumina sequencing. These PCR conditions are described in Appendix 6.3. The third round of PCR was used to add the Illumina indexes (Illumina, 2016), with the conditions described in Appendix 6.4.

The product was visualized in a 1.6× agarose gel stained with ethidium bromide. All samples that showed amplification near the appropriate fragment length were examined using a bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) for quantitation and to inform magnetic bead clean-up methods. Clean-up was conducted using the Agencourt AMPure XP beads (Bekman Coulter, Inc. Indianapolis, IN) following the manufacturer's protocol, with a 1:1 ratio of bead to product. The product was quantified with qPCR and pooled accordingly to achieve a 4nM final concentration for sequencing on an Illumina MiSeq using manufacturer protocols (Illumina, San Diego, CA) at the Griffith University DNA Sequencing Facility.

6.3.7 Bioinformatics

Sequence reads were demultiplexed by sample using the MiSeq native bcl2fastq2 command. Reads were merged using PEAR v0.9.10 (Zhang et al., 2013), then quality filtered, de-replicated, checked for chimeras and clustered at 97% using UPARSE version 8.0.1623 (Edgar, 2013). Taxonomic assignment was performed using a naïve Bayesian classifier (Wang et al., 2007) against a custom database. The latter database was constructed by *in silico* PCR (Bellemain et al., 2010, Ficetola et al., 2010) using the primers from the present study to amplify homologous DNA regions from the NCBI nt database (accessed June 2016) allowing up to three base pair mismatches per primer.

The ten sequences with the highest number of reads were run in BLAST (Benson et al., 2012) and the top ten hits were imported to Geneious (version 8.1.1; Biomatters, Ltd) to test for alignment with primers.

6.4 Results

The taxonomic class diversity and proportions of eDNA found from sequencing were sorted by abundance of DNA reads (Figure 6.2). Prior to sequencing, the bioanalyzer results showed that the majority of product after bead clean-up was about 255bp, but the target length was 225bp. After the sequencing run and bioinformatic filtering, no sequenced DNA was identified as belonging to fish except for some sequences that had assignments with low statistical support. OTUs assigned to fish were checked with BLAST (NCBI) and the descriptions returned were primarily bacterial with low support for assignments at the species level (Table 6.3). Therefore, these results were inconclusive. Primarily bacterial DNA was identified with sequencing.

The overwhelming majority of reads were assigned to the bacterial class Anaerolineae (8,982,998 reads, average support 76.9, SD±17.8). The second highest proportion of reads were assigned to Mammalia (280,210 reads, average support 91.2, SD±9.3), followed by the bacterial class Deinococci (132,569 reads, average support 37.1, SD±17.9). A smaller proportion of reads were assigned, with low support, to the classes Aves (63,399 reads, average support 29.5, SD±14.3), Actinopteri (41,106 reads, average support 39.7, SD ±8.3), Amphibia (29,063 reads, average support 27.3, SD±9.3), and the bacteria Thermotogae (1,535 reads, average support 25.3, SD±8.9) (Figure 6.2).

Sequencing identified 31 species of fish in all samples combined, with low support below the phylum level for matches in GenBank. A list of species names is shown in Table 6.3, along with support levels for taxonomic assignment at the species level.

Based on visualization with gel electrophoresis, eDNA was successfully amplified at the length expected based on target length of the fish primers in a few layers of surface sediment, but further experiments would be needed in order to confirm if these bands contained fish eDNA. Several non-fish vertebrates were identified, including amphibians, birds, bat, whale, and order Squamata (scaled reptile), but these sequences likely represent errors.

For the ten most abundant OTUs, the top ten hits in BLAST nearly all showed alignment with the forward end primer (Teleo_F), with three mismatches near the 3' end of the marker. An example from the alignment for the OTU with the highest number of reads is shown in Figure 6.3.

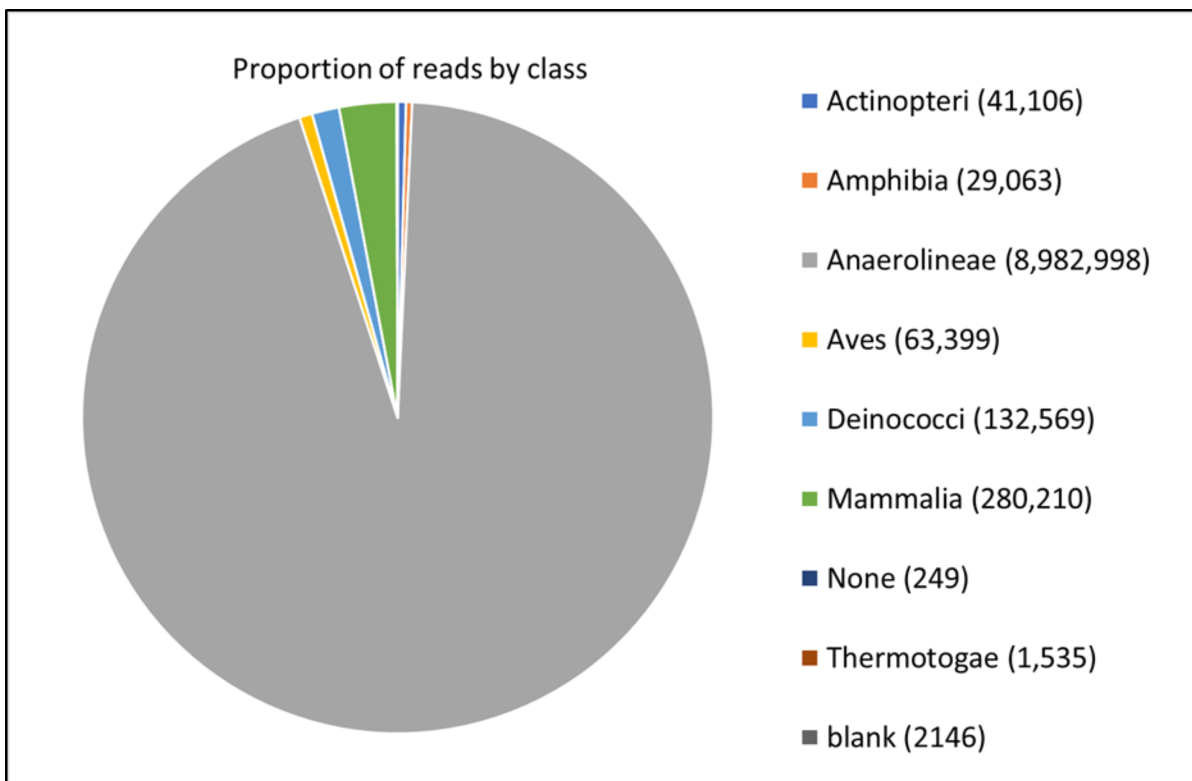

Figure 6.2: Proportion of reads assigned to different classes. Read counts for each class are shown in parentheses. "None" category did not achieve class-level assignments, and "blank" category received no assignments.

Table 6.3: List of results from top hits in BLAST runs

Fish Species	Species-level support	BLAST description	E value	% Identity	Query cover(%)
Antennarius striatus	14	None			
Archamia bleekeri	10	Bacterium clone	2e-42	100	97
Benthosema pterotum	18	Archaeon clone	4e-34	95	97
Boulengerella maculata	13	Bacterium clone	2e-47	99	97
Calophysus macropterus	10	None			
Cottus dzungaricus	10	Bacterium clone	9e-46	98	97
Cynodonichthys weberi	11	Bacterium clone	2e-43	96	97
Cynoglossus abbreviatus	17	None			
Diretmoides veriginae	16	None			
Ginsburgellus novemlineatus	11	Bacterium clone	9e-46	98	97
Harttia longipinna	11	Bacterium clone	4e-40	94	97
Hypselecara temporalis	10	Bacterium clone	9e-46	99	97
Lipolagus ochotensis	10	None			
Lophogobius cristulatus	13	None			
Macquaria australasica	16	None			
Mystus mysticetus	11	Archaeon clone	8e-41	99	97
Opsaridium microlepis	24	None			
Ostorhinchus hoevenii	12	Bacterium clone	6e-28	89	97
Oxymonacanthus longirostris	10	Bacterium isolate	9e-41	96	97
Oxyurichthys formosanus	16	Archaeon clone	8e-36	96	97
Pardachirus pavoninus	33	Bacterium clone	2e-43	97	97
Pempheris schwenkii	13	Bacterium	1e-30	90	97
Pteragogus amboinensis	11	Bacterium clone	3e-46	99	97
Ptereleotris zebra	16	Archaeon clone	5e-38	98	96
Sardinella longiceps	16	Microorganism clone	7e-47	99	97
Scriptaphyosemion guignardi	16	Proteobacterium clone	2e-47	100	97
Selar crumenophthalmus	15	Archaeon clone	3e-35	96	97
Sigmops gracilis	20	Archaeon clone	0.001	100	28
Solivomer arenidens	12	Bacterium clone	3e-35	93	97
Tomeurus gracilis	15	None			
Zeus capensis	13	Bacterium clone	4e-39	95	97

```
                                        1        10        20        30        40        50        60        70        80        90       100
1. HM748685_Bacterium_Ellin6537         ACACCGCCCGTCACGTCATGGGAGCTGGCAACACCTGAAGCCGGTGCCCGAACC--GCAAGGACGGAGCCGTCGAGGGTGGGGCTGGTGACTGGGACGAA
2. KM461971_Anaerolineae_bacterium      ........................T....................AT..T....--.......G.........A.A.......T..A.A...........
3. KM018331_Chloroflexi_bacterium       ........................T....................AT..T....--.......G.........A.A.......T..A.A...........
4. KF030201_Chloroflexi_bacterium       ........................T....................AT..T....--.......G.........A.A.......T..A.A...........
5. JN713473_Chloroflexi_bacterium       ........................T....................AT..T....--.......G.........A.A.......T..A.A...........
6. NR_133813_Pelolinea_submarina        .....................A.......................AT..T....--.......G...................TC.............
7. NR_040971_Leptolinea_tardivitalis    .............................................AT..T....--.......G...................TC.............
8. NR_117865_Thermanaerothrix_daxensis  .............................................AG..T....CGCA.G...G..C......N.A........................
9. NR_040972_Levilinea_saccharolytica   .............................................AT..T....CGCA.G...G...........A.......T................
10. LT859958_Anaerolineaceae_bacterium  .....................A..........G............AT..T....CTCACG...G.........T.A.......T................
11. Teleo_F_L1848                       ..............TCT
```

Figure 6.3: Alignment of the Teleo_F PCR primer (line 11 of alignment) with the top ten hits from a BLAST of the OTU with the highest number of reads. Dots represent nucleotides that are shared with the reference sequence (line 1 of alignment). The three bases at the 3' terminal of the primer do not match the bacterial sequences that amplified (column 15, see box).

6.5 Discussion

The failure of the current study to identify fish DNA sequences may be related to an overwhelming abundance of bacterial DNA in the samples and/or the metabarcoding approach being reliant on generic fish primers. That is, it is likely that the relatively high quantity of bacterial DNA in samples swamped out the signals from expected low proportions of fish eDNA. The ten most abundant OTUs were run in BLAST and the top ten hits were imported to Geneious (version 8.1.1; Biomatters, Ltd) to test for alignment with primers. Nearly all hits showed alignment with the forward end primer (Teleo_F), with three mismatches near the 3' end of the marker (Figure 6.3). Adding phosphorothioate links to the 3' ends of the primer to prevent degradation could provide a partial remedy to the problem (Zon and Geiser, 1991), but might not overcome the challenge of the target being in too low concentration relative to background eDNA. It is also possible that eDNA of fish was highly degraded, and fragment lengths were too short for primers which targeted ~100bp fragment of the 12S region. More experiments are required.

Sequencing results were overwhelmingly identified as bacterial despite the use of PCR to target and amplify fish eDNA. Sequencing depth has been shown to be a critical factor in the ability to detect rare species and make ecological inferences from metabarcoding data (Smith and Peay, 2014). Some fish were superficially identified in the sequenced eDNA, but the proportion of sequences or levels of support for species-level matches with reference sequences from GenBank were too low for them to be considered valid assignments (Table 6.3). The fact that non-fish vertebrates were also identified may indicate that the primers were not specific enough. It is challenging to design a single universal primer set that will amplify all bony fish species, but that excludes other vertebrates for the short fragments that are found in eDNA samples. Indeed, fish share a common ancestor with nearly all vertebrates, and therefore share

large portions of their genomes. There are published studies which successfully identified ancient fish eDNA from sediment cores (Matisoo-Smith et al., 2008, Pedersen et al., 2016), though the samples were collected in colder climates which are more amenable to the preservation of eDNA over long time periods and their laboratory protocols were different.

The overwhelming majority of reads were assigned to the bacterial class Anaerolineae. This group is primarily made up of filamentous anaerobes, indicating this DNA came from live and actively metabolizing bacteria that consume organic matter such as eDNA particles. The second highest proportion of reads were assigned to class Mammalia, and identified primarily as human DNA. The third highest proportion of reads were assigned to the bacterial class Deinococci, which is made up of highly resilient bacteria with thick cell walls that can withstand extremes in environmental conditions (Griffiths and Gupta, 2007). Figure 6.2 shows that a much smaller proportion of reads were assigned to birds, fishes (Actinopteri), and amphibians. Levels of support for chordates at the class level ranged from an average of 27.4 for Amphibia to 91.2 for Mammalia. The Mammalia results could be from endogenous sources or more likely from contamination during processing of the samples. It is unlikely that the assignments for other vertebrates were accurate, given the low levels of support. When some of the sequences identified as Mammalia were run in BLAST, the description returned was *Homo sapiens* with query cover of 100% and identity of 99% (E value 5e-37).

Pedersen et al. (2016) used cores collected in lakes of the Peace River Basin of western mid-Canada to successfully identify ancient eDNA of animals and plants. They used shotgun sequencing of the full metagenome present in the samples, and detected ancient eDNA of the fish *Esox* (pike) in core layers dated as far back as $\sim$ 11.7 thousand years before present. However, faunal remains of fish in a nearby cave were dated to

12.4 thousand years before present. The inability to detect fish in layers dating to 12.4 thousand years before present may be the result of eDNA degradation, however their ability to identify sequences in sediment from over 11 thousand years ago is an indication of high rates of DNA preservation likely due to cold temperatures. This could be one explanation for their ability to detect ancient fish eDNA while the current study did not. Matisoo-Smith et al. (2008) sampled a sediment core collected in a New Zealand lake and were able to identify DNA that was very similar to a native fish species, common bully (*Gobiomorphus cotidianus*). They used a 12S primer designed for a native bat species and direct sequencing. They validated this finding by amplifying and sequencing DNA extracted from tissue samples from a modern specimen of common bully.

If the current study was repeated, use of alternative laboratory methods may enable identification of fish eDNA from sediment layers. Nested PCR with primers designed for the CO1 region of Australian lungfish (*Neoceratodus forsteri*) were tested on a small number of samples and showed bands of appropriate length in agarose gels. These primers were designed to target shorter fragments than the Teleo primers. For future studies attempting to amplify ancient fish eDNA from sediment cores, it would be advisable to use alternative methods to those described in this chapter. One alternative which may prove more effective is hybridization enrichment (Mamanova et al., 2010, Hofreiter et al., 2015). Another alternative is to design a several primer sets that are species-specific and target very short fragments with a nested design, based on a fish species that is very common in the Brisbane River. This could reduce the amplification of non-target DNA and reduce the swamping effect that may have occurred in sequencing for this study, and also amplify very short (i.e. highly degraded) fragments of DNA.

It is likely that the samples used in this study were subject to eDNA degradation rates much higher than those for other studies of ancient eDNA in sediments, given the subtropical climate of Moreton Bay and the Brisbane River catchment. A global study of DNA degradation rates based on temperature regimes showed that Australia had one of the lowest expected DNA survival rates (Hofreiter et al., 2015). Therefore, it could be the case that fish eDNA is not preserved in subtropical sediments for long time periods.

The endosymbiotic theory states that mitochondria are descendent from bacteria that were assimilated into another cell (Sagan, 1967). Metabarcoding results for this chapter may have been dominated by bacteria because the markers may have consistently amplified a homolog of the 12S gene. The primers used in the current study targeted a mitochondrial region (12S), and did not appear to amplify bacteria haphazardly, but instead rather selectively. The primers amplified certain clades of bacteria, but not others that would be expected to be present in the sediment. Therefore, the results presented here do not represent relative abundance of bacterial communities, because the primers may have preferentially amplified certain bacteria over others. This finding has implications for the design of eDNA experiments that target extremely low copy number template. Most eDNA studies target the mitochondria rather than nuclear loci (Deiner et al., 2017), because mitochondrial genomes are more abundant than nuclear genes and therefore have higher detection rates. However, the results of this chapter show that care needs to be taken when designing molecular studies that target genomic regions that have microbial homologs.

While a few studies of ancient eDNA in sediments have successfully identified fish eDNA, these are rare in the published literature and to our knowledge none have been conducted on sediment cores collected from a subtropical environment. Further work and laboratory experiments are needed to determine approaches that may enable

the identification of ancient fish eDNA from sediment cores collected in warmer climates. Understanding the impacts of past land-use change by humans or influence of variations in climate regimes on biotic communities is relevant for predicting future scenarios and developing improved management strategies.

CHAPTER 7: GENERAL DISCUSSION AND SYNTHESIS

Rapidly advancing eDNA analysis methods for the study of aquatic macroorganisms present a great deal of promise; however, there are still many limitations to be resolved. Given the high rates of decline for freshwater biodiversity, any tool with the potential to provide improved understanding of spatiotemporal distributions of threatened taxa is worth pursuing. A review of the eDNA literature related to fish (Chapter 2) showed that while this technique is under development, but has been successfully applied in a variety of contexts. In this thesis I attempted to improve understanding of the distribution of a critically endangered species in a large tropical river in order to inform conservation strategies (Chapter 3). The process of applying eDNA techniques in such a complex environment clarified the need for further work to refine the method. I aimed to test whether choice of generic fish primers influences information on taxa composition, in order to understand the extent to which this tool can provide broad information on the distribution and composition of fish communities (Chapter 4). The finding that different primers returned varying taxa composition estimates with eDNA metabarcoding (Chapter 4), led to further work to test whether the heterogeneity of rivers and the choice of sampling location within the river channel influence results (Chapter 5). Finally, in order to understand the potential of sediment to preserve eDNA and provide a chronological record of past fish community composition, I tested the technique on sediment cores collected in a subtropical embayment at the outflow of a large river (Chapter 6). This body of work contributes to the rapidly growing knowledge on the potential and limitations of eDNA tools for the study of aquatic macroorganisms. I pursued this work under the broader goal of gaining access to knowledge on rare and threatened species that could enhance field survey methods and provide information to support more effective conservation strategies. eDNA tools are not at the point of

development where they should be seen as a replacement for conventional, observation-based, field survey methods for taxa richness and assemblage data. These tools should be viewed as complementary to conventional methods, and currently are best applied for catchment-level analysis of diversity, for rapid "first pass" surveys to identify and prioritize locations for more in-depth studies using conventional methods, and for locations or taxa that are not accessible with conventional survey gears.

Chapter 2 examined the benefits and limitations of the application of eDNA metabarcoding as an ecological survey tool compared to conventional survey methods. The review found that a large proportion of the publications have focused on testing the sensitivity of the methods. Within the category of methods development, publications primarily examined environmental and biological factors that influence rates of eDNA shedding, persistence, degradation, transport, retention, and detection probabilities (Figure 2.2). The first papers that compared results from conventional survey methods to eDNA survey methods from field experiments were published in 2015, and the number of such studies grew in 2016. The results of the review characterize the current status of eDNA methods as being in the 'developmental phase'. The growth in the number of publications that used conventional methods in tandem with eDNA methods indicates the need to validate the molecular methods with observational data, given the uncertainties and potential for biases associated with molecular approaches. This chapter also contextualized the use of molecular tools for conservation application given the limitations on currently available data for rare freshwater species and large tropical rivers.

In Chapter 3 I tested the utility of the eDNA method for determining the presence of a rare species. This is the first time that eDNA methods have been applied in such a large river in a tropical environment. Mekong giant catfish DNA was detected in one out

of the six sampling sites. The location where it was identified is the only known spawning habitat for *P. gigas*. This location is upstream of the first mainstream hydroelectric dam being built in the portion of the river downstream from China, known as the Lower Mekong Basin. Whether the giant catfish will persist in the wild once this dam is built is yet to be seen, but future eDNA surveys upstream and downstream of the dam may help to answer this question. High water volumes, depths, and turbidity in the Mekong River present extreme logistical challenges for the deployment of conventional survey methods. These conditions also complicate the use of eDNA methods. However, given the challenging environmental conditions, it is not feasible to apply conventional methods to study the distribution of *P. gigas* in the wild. The only other tool that has been used to survey the deep pools which giant catfish are expected to inhabit is sonar (Viravong et al., 2006). However, sonar only provides information on biomass and not on species identity. Given that it was not possible to use alternative observation-based methods in the same locations as eDNA sampling, it is not possible to assess whether any lack of detection in the five other sites can be attributed to true absence of the species in those sites or failure of the eDNA tool to detect them.

In Chapter 4 three different sets of bony-fish-specific PCR primers were used to conduct eDNA metabarcoding of water samples collected at four sites in a subtropical river. This is the first time that multiple sets of primers targeting fishes were used simultaneously on the same samples to test the assumption that a single generic fish primer can return unbiased taxa composition data. The results of metabarcoding were also compared to richness and taxa composition results returned from conventional sampling (seining) conducted simultaneously with eDNA sampling, a list of species known to occur based on past surveys with conventional methods (Pusey et al., 2004), and species occurrence data from modeling. Some previously published studies have

used clade-specific primers and compared these results with data from conventional sampling in order to validate the claim that eDNA metabarcoding provides more information on species composition (Thomsen et al., 2012b, Valentini et al., 2016).

However, the tests performed in this thesis show that results from a single primer set may not be reliable. Additionally, most studies compare eDNA results with those obtained from a single type of conventional sampling gear. A recent study which used American Fisheries Society (AFS) standard sampling methods of gillnetting and electrofishing in tandem with eDNA collection showed that eDNA did not always return results which agreed with those found using AFS standard methods (Perez et al., 2017). In some of the sites where fish were detected with standard methods, species-specific primers failed to detect them. The authors concluded that eDNA methodology needs further refinement before it can be included in standard sampling efforts. Their results concur with results described in Chapter 4 of this thesis. For example there were two genera (*Craterocephalus*, *Oreochromis*) detected with seine netting that were not identified though eDNA metabarcoding. However, two genera not detected with seining were identified through eDNA and deemed likely to occur based on modeling and historic records (*Anguilla*, *Mogurnda*) (Chapter 4, Appendix 3). The failure to identify two genera that were observed at the time of eDNA sample collection can likely be attributed to problems in the molecular workflow. These include sequencing or primer bias, stringency of bioinformatics filtering, availability of reference genomes in libraries, the possibility that eDNA is not equally distributed at a sampling location, or that the choice of sampling location within the river cross-section influences the collection of representative samples.

In Chapter 5 I tested the effect of sampling location within the river channel on eDNA metabarcoding results for richness and taxonomic assemblage. This is the first

analysis in a natural river of the influence of different hydraulic units and the position of sampling across the river channel on eDNA metabarcoding results. This chapter analyzed the role that physical variability in river channels may play in causing variance in richness and taxa assemblage estimates based on metabarcoding with generic fish and eukaryote primers. True replicate samples were collected at three different points across the river channel in both a pool and a riffle, at three different times of day. The factors of hydraulic unit, channel position within a hydraulic unit, and time of collection for a given channel position within a hydraulic unit were found to have significant effects on taxa assemblage identified with the eukaryote primers. For the fish primers, none of the factors tested in the model showed significant effects. Few of the replicate samples returned identical results for richness or taxa assemblage, and there was a clear difference in taxa assemblage from samples collected in the pool compared to those collected in the riffle. This could partially be due to a true difference in composition between these hydraulic units, but given that the riffle was immediately downstream it would be expected to contain the eDNA that was contained in the pool, as eDNA drifts downstream and has been found to be detectable downstream between 239m (Jane et al., 2015) and up to 12km in some cases (Deiner and Altermatt, 2014). The measurement of isovels and surface velocity showed high variation between the pool and the riffle (Chapter 5, Figure 2). It is likely that velocity is an important factor in eDNA detection rates, and can influence the ability to collect a sample which contains eDNA that is representative of the full suite of taxa present in any given site. For the majority of eDNA studies in rivers, samples were either collected at a single point location or the precise location of sample collection was not reported. To my knowledge Chapter 5 represents the only study that compared eDNA detection among hydraulic units or from samples taken at different points of a cross-section in a natural river. Even

within replicate samples collected at the same point locations there was high variation in eDNA metabarcoding results, particularly with the clade-specific primer set. The results of chapter 5 seem to indicate that the location and timing of sample collection will influence estimates of richness derived from them. These findings support the conclusion that there is still a great deal we do not understand about the "ecology" of eDNA (Barnes and Turner, 2015).

Chapter 6 tested the potential for sediment cores collected in a marine embayment at the outflow of a subtropical river to preserve a chronology of freshwater taxon presence and assemblage over long time periods. This is the first time that sediment collected in a marine bay and sourced from a subtropical river has been examined for the presence of eDNA from freshwater fish. It was hypothesized that increases in sedimentation rates associated with land use changes from European settlement of the Brisbane River catchment may be associated with shifts in assemblages of river fishes. After testing several extraction protocols, eDNA from sediment layers dated to both pre and post-European settlement was amplified with generic fish primers. A high proportion of OTUs identified with sequencing were assigned to bacterial reference sequences with low support. Therefore, these results were inconclusive. When amplified with separate primers designed specifically for Australian lungfish (*Neocerotodus forsteri*), several layers of surface sediment showed bands at the expected length when visualized with gel electrophoresis. The high proportion of background bacterial DNA in samples relative to eDNA of fishes may be an explanation for these inconclusive results. Also, the forward end of the Teleo primers was found to align with bacterial sequences when tested in BLAST, with three mismatches at the 3' end. Further work on these samples may be more successful if an alternate molecular workflow was applied. Adding phosphorothioate links to the 3' end of the primer to

prevent degradation might offer a partial remedy to the problem (Zon and Geiser, 1991). Additionally, the use of capture hybridization could prove more effective than traditional PCR and this method has great potential for eDNA applications (Jones and Good, 2016). Capture hybridization is a laboratory technique which enables detection of eDNA fragments without the use of PCR amplification and which can target shorter amplicon lengths than PCR. This technique has primarily been used for studies of ancient DNA (Ávila-Arcos et al., 2011, Mason et al., 2011, Hofreiter et al., 2015). Further application and development of capture hybridization can advance eDNA studies, and it will likely be applied more widely in the near future.

7.1 Next steps for advancing conservation through eDNA tools

In some cases, eDNA metabarcoding can provide data for species and systems that have been too difficult or costly to survey in the past using observation-based (conventional) methods. For example, species that occur in very low densities over extensive ranges and in river systems with high discharge. Given certain conditions, eDNA methods may help to overcome some of the limitations of conventional sampling, especially in environments presenting logistical challenges to physical capture or observation of organisms. Some of the conditions in which eDNA methods are most likely to provide reliable detections include: low flows, lentic waters, and cold temperatures. The eDNA metabarcoding approach presents some specific challenges due to the high levels of DNA degradation after it is shed and relatively low abundance of target molecules in samples when studying macroorganisms. For the study of highly rare aquatic species, the method requires further refinement to support broad-scale application.

To validate richness and diversity estimates based on metabarcoding data, some studies use control mixtures of genomic DNA in known concentrations extracted from

different species that were collected from, or known to occur in, the study area (Yu et al., 2012, Ji et al., 2013, Leray, 2015). Throughout this thesis, such an approach is distinguished from eDNA metabarcoding as "DNA metabarcoding", following the definitions established by Taberlet et al. (2012a). Conclusions from studies based on metabarcoding of high quality/quantity genomic DNA cannot necessarily be applied to the study of eDNA. The low quality/small size of eDNA fragments, influence of environmental factors on degradation rates (Barnes et al., 2014), adsorption by clay minerals (Pietramellara et al., 2007), as well as transport and settling rates (Deiner and Altermatt, 2014, Jane et al., 2015, Shogren et al., 2017) of eDNA molecules in natural water bodies introduces unique challenges.

To address PCR and sequencing biases, metabarcoding results are subjected to quality filtering before statistical analysis. Thresholds for read quality are set, and remaining reads are clustered according to similarity using published algorithms. The clustered sequences are then compared to reference genome sequences from online or custom-built libraries to assign taxonomy. Selection of clustering thresholds and algorithms influences estimates of richness and beta diversity. Richness estimates have been shown to vary by three orders of magnitude depending on which combination of these analytic methods is applied (Flynn et al., 2015).

Based on the results of Chapter 4, the Teleo primers produced the highest richness estimates compared to the other two primer sets used, but this value dropped substantially when the clustering threshold was reduced. This indicates that the Teleo primers were able to detect a higher number of unique sequences, but it is unclear whether these sequences all represent true species. This could indicate that the locus is highly variable or not variable at all, but it is more likely caused by the lack of adequate reference sequence information being available in the NCBI database for Mary River fish

species. In some cases, it may not be possible to distinguish true sequences from artifacts without matching reference genomes. Given the rapidly dropping price for genetic sequencing associated with high-throughput technology, more effort should be dedicated to sequencing for reference libraries.

Type I errors are less likely to occur than Type II errors with eDNA metabarcoding, but false detections are a challenge that needs further work. Some potential causes of false detections include: sample contamination, detection of dead organisms, resuspension of sediment that has adsorbed eDNA, primers that are not specific enough, PCR and sequencing errors, and misidentifications in reference libraries (Ficetola et al., 2015, Roussel et al., 2015, Wilcox et al., 2013). In Chapter 4 of this study the clade-specific primers and GenBank assignment returned an incorrect identification at the species level for *Leiopotherapon plumbeus*, which is not known to occur in the Mary River. However, the closely related species *Leiopotherapon unicolor* does occur. This example illustrates the importance of using alternate sources of biodiversity data in tandem with eDNA work, and the challenge of designing primers that can amplify all species within a diverse clade while also identifying sequences down to the species level. The high sensitivity of eDNA protocols, some of which can detect a single copy of target DNA (Wilcox et al., 2016), translates to a high susceptibility for sample contamination. False positive detections can occur from contamination in the field or the lab. Rigorous protocols for equipment sterilization and collections in the field have been published by Carim et al. (2015).

Conservation actions will benefit further from eDNA analysis when the tool advances to the point where it can be used to reliably estimate abundance or biomass of target organisms. Perez et al. (2017) did not find any relationships among the catch or biomass per-unit-effort found with conventional sampling and the fish eDNA copy

numbers for samples collected in two different seasons in a reservoir. They did find that copy number varied with season, and therefore recommended that eDNA sample collection methods include a temporal component. Elbrecht and Leese (2015) found that the relative abundance of DNA from mock communities was skewed after amplification because of primer bias from mismatches with the target. Some studies have shown a link between eDNA concentration and biomass (*sensu* Kelly et al. (2014)), though based on a review of the literature there is too much variability in eDNA results to use them as indicators of biomass at the moment (*sensu* Lacoursière-Roussel et al. (2016b)).

eDNA methods are still in the early stages of development, and with time will likely advance to provide higher detection sensitivity. However, accurate interpretation of results will require further understanding of factors which control error rates. Some parallels can be drawn between eDNA methods and diagnostic testing for biomedical applications. Biomedical diagnostics publications have been shown to lack an assessment of what controls the rate of false discoveries (error rate) (Colquhoun, 2014). Significance levels assigned to results are only accurate if all the assumptions made in an experiment are correct (Colquhoun, 2014). Many eDNA studies include assumptions which we are learning to be false. As discussed in Chapter 5, collecting eDNA samples from only the river bank is unlikely to provide a representative sample, yet many eDNA studies assume eDNA is homogenous in a water body. The incorporation of lessons learned in other relevant disciplines should be a priority for future eDNA studies. Despite the need for further development, eDNA analysis has been shown to currently enable collection of data that are useful for conservation applications.

Given the dire outlook for freshwater species, there is not time to perfect eDNA metabarcoding techniques before they are applied in a conservation context. However,

given the current limitations, eDNA analysis may be best suited for use in combination with conventional methods. All field survey methods have limitations. eDNA analysis can be used to rapidly and affordably identify locations that should be prioritized for more rigorous sampling with conventional methods. Metabarcoding data can expand the spatiotemporal coverage and improve the sensitivity of ecological models that predict the influence of environmental variables on diversity. This is especially relevant for aquatic species, which are typically highly mobile and under-surveyed. However; if high spatial resolution of occupancy data is the goal of research, downstream drift of eDNA complicates application of this method in river environments. Field experiments and modeling have shown detection probability for a restricted macroinvertebrate did not drop below 95% until 50km downstream (Deiner and Altermatt, 2014). Distance of downstream detection for trout has been shown to be influenced by flow levels and deposition of leaf litter which contains PCR inhibitors (Jane et al., 2015). For research in rivers, eDNA metabarcoding may currently be most useful for characterizing diversity at larger scales, such as entire catchments (Deiner et al., 2016), or for distinct segments within a dendritic network (i.e. comparing tributaries to mainstems or to each other).

7.2 Broader implications for conservation

Currently, freshwater species face the highest rates of decline and one third of freshwater vertebrates are classified as threatened on the Red List (IUCN, 2018). The grim state of freshwater biodiversity cannot be overstated. Given this crisis, it is warranted to test any and all potential tools that may provide data that will support more effective conservation actions. River basins around the globe have been transformed in the developed world, and are currently undergoing extensive change in the developing world. Human needs and demand for freshwater often appear to conflict with biodiversity preservation. Enhanced management based on information gathered

rapidly and affordably can alleviate some conflicts, and novel solutions may be discovered. We cannot manage threatened species when we lack knowledge of their spatiotemporal distributions at scales which are relevant for biological, social, and political needs.

Conventional observation-based tools to assess spatiotemporal distributions are limited in their scope of application given the required high levels of time and financial investment to deploy them. When considering global coverage, surveys are highly biased towards the northern hemisphere and temperate aquatic systems. However, large tropical rivers host the highest species richness and highest number of threatened species, while the current level of knowledge is most limited. The physical conditions, such as high habitat heterogeneity, high discharge levels, and high seasonal variation in discharge; deep waters, high levels of turbidity, and presence of large rapids and amounts of debris severely limit the extent to which observation-based methods can be applied. The expansive geographic and trans-boundary extent of many large tropical rivers make it all the more challenging to conduct representative surveys and to gain access to sampling areas in different political jurisdictions. Given these challenges, it is worthwhile to pursue alternative techniques like eDNA analysis to study such systems, despite the current limitations of these tools.

8: APPENDICES

- Appendix numbering is defined by associated chapters

Appendix 4.1: Standard PCR Conditions

Standard PCR was conducted under the following conditions: each reaction contained a volume of 30µL, with final concentrations of 1x supplied buffer (New England Biolabs, Inc., Ipswich, MA, USA), 0.2mM dNTPs, 0.5mg/mL BSA (New England Biolabs, Inc., Ipswich, MA, USA), 0.05µM of each primer, and 0.1 units of Q5 hot start polymerase (10units/mL, New England BioLabs, Inc., Ipswich, MA, USA). Each reaction contained 1µL of eDNA extract. Thermal cycling was at 98°C for 30 sec (x1 cycle); 98°C for 10 sec, primer-specific T_m (Table 1) for 5 sec (Table 1), and 72°C for 5 sec (x20); final extension at 72°C for 30 sec. Three replicates were performed for each sample, then products were pooled and diluted 1:25 with molecular grade water.

Appendix 4.2: PCR conditions for Illumina adapters

Illumina adapters were added under the following conditions: each reaction contained a volume of 22 µL, with final concentrations of 1x supplied buffer, 0.2 mM dNTPs, 1 mg/mL BSA (New England BioLabs, Inc., Ipswich, MA, USA), 0.02 µM of each adapter, and 0.25 units of Q5 hot start polymerase (New England BioLabs, Inc., Ipswich, MA, USA). 5 µL of diluted product from the first round of PCR was added. Thermal cycling was at 98°C for 30 sec (x1 cycle); 98°C for 10 sec, 59 °C for 5 sec, and 72°C for 5 sec (x30); 72°C for 30 sec (x1). The product was subjected to electrophoresis in a 2% agarose with 0.5xTBE, stained in 50 ng/ml ethidium bromide in 0.5xTBE, and visualized over UV light.

Appendix 4.3

List of fish genera positively identified by different sources of richness data for the Mary River, Queensland, Australia. For eDNA results “x” indicates identification at genus level, “o” indicates identification at the family or order level. Green shading indicates high confidence in eDNA results, yellow shading indicates moderate confidence given alternate data sources, and orange shading indicates low confidence in eDNA results (i.e. further survey work required to validate eDNA results, or eDNA failed to detect species observed with seining). Predictive modeling probability was scored as positive if genus had a probability ≥0.6 at one or more sites.

Genus Name	eDNA	Seining	Predictive modeling (prob≥0.6)	Reference list (Pusey et al. 2004)
Ambassis	x	x	x	x
Anguilla	x		x	x
Craterocephalus		x	x	x
Gambusia	x	x		x
Glossamia	x	x		x
Hypseleotris	x	x	x	x
Leiopotherapon	x			x
Maccullochella	x			x
Melanotaenia	x	x	x	x
Mogurnda	x		x	x
Mugil	x			x
Nematalosa	x			x
Neoceratodus	x			x
Notesthes			x	x
Oreochromis		x		x
Philypnodon			x	x
Pseudomugil	x	x	x	x
Rhadinocentrus				x
Retropinna	x	x	x	x
Tandanus	o		x	X
Xiphophorus	x			x

Appendix 6.1: Methods and results for trials of three ratios of extraction buffer to sediment

We selected sediment samples from layers that covered a high range of dates (MB2L_20-30, MB2L_180-190, MB2L_330-340) to evaluate whether different ratios of binding buffer worked more effectively for highly degraded eDNA in very old samples or if results were uniform regardless of sample age.

For each of three samples we tried three different ratios of binding buffer:

- 1a = 300uL supernatant (1/2 of total from extraction) + 2x volume of binding buffer (compared to manufacturer protocol)
- 1b = 200uL of supernatant (1/3 of total from extraction) → 3x volume of binding buffer (compared to manufacturer protocol)
- 1c = 100uL of supernatant (1/6 of total from extraction) → 6x volume of binding buffer (compared to manufacturer protocol)

Appendix 6.2: Standard PCR conditions, Round 1

Standard PCR was conducted under the following conditions: each reaction contained a volume of 15μL, with final concentrations of 1x supplied buffer, 0.2mM dNTPs, , 0.1μM of each primer (Teleo_F and Teleo_R), and 0.02 units/μL of KAPA HiFi polymerase (KAPA Biosystems, Boston, Massachusetts, United States). Each reaction contained 1μL of eDNA extract. Thermal cycling was at 95°C for 3 min. (x1 cycle); 95°C for 20 sec (x25), 56°C for 8 sec.; 72°C for 3 sec.; and 72°C for 30 sec (x1).

Appendix 6.3: Standard PCR conditions, Round 2

Standard PCR was conducted under the following conditions: each reaction contained a volume of 15μL, with final concentrations of 0.8x supplied buffer, 0.2mM dNTPs, 0.1μM of each primer (Teleo_F and 12SiiSHRT_R), and 0.02 units/μL of KAPA HiFi polymerase (KAPA Biosystems, Boston, Massachusetts, United States). Each reaction contained 3μL of PCR product. Thermal cycling was at 95°C for 30 sec. (x1 cycle); 95°C for 10 sec (x25), 70°C for 4 sec.

Appendix 6.4: Illumina indexing conditions

Indexing was conducted under the following conditions: each reaction contained a volume of 22μL, with final concentrations of 1x supplied buffer, 0.2mM dNTPs, 0.1μM of indexing primer, and 0.02 units/μL of KAPA HiFi polymerase (KAPA Biosystems, Boston, Massachusetts, United States). Each reaction contained 0.1μL of product. Thermal cycling was at 95°C for 3 min. (x1 cycle); 95°C for 20 sec (x30), 64°C for 8 sec.; 72°C for 3 sec.; and 72°C for 30 sec. (x1).

REFERENCES

ADAMS, R. I., AMEND, A. S., TAYLOR, J. W. & BRUNS, T. D. 2013. A unique signal distorts the perception of species richness and composition in high-throughput sequencing surveys of microbial communities: a case study of fungi in indoor dust. *Microb Ecol,* 66**,** 735-41.

AITKEN, M. J. 1998. *Introduction to optical dating: the dating of Quaternary sediments by the use of photon-stimulated luminescence*, Clarendon Press.

ALLAN, J. D., ABELL, R., HOGAN, Z., REVENGA, C., TAYLOR, B. W., WELCOMME, R. L. & WINEMILLER, K. 2005. Overfishing of Inland Waters. *BioScience,* 55**,** 1041-1051.

ALLEN, G. R., MIDGLEY, S. H. & ALLEN, M. 2003. *Field Guide to Freshwater Fishes of Australia, 2nd edn.,* Perth, Western Australian Museum.

ALLOUCHE, O., TSOAR, A. & KADMON, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology,* 43**,** 1223-1232.

AMOS, W., WHITEHEAD, H., FERRARI, M. J., GLOCKNER-FERRARI, D. A., PAYNE, R. & GORDON, J. 1992. RESTRICTABLE DNA FROM SLOUGHED CETACEAN SKIN; ITS POTENTIAL FOR USE IN POPULATION ANALYSIS. *Marine Mammal Science,* 8**,** 275-283.

ANDERSON, M., GORLEY, R. & CLARKE, K. 2008. PERMANOVA+ for PRIMER: guide to software and statistical methods. 2008. *Plymouth, UK*.

ANDERSON, M. J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology,* 26**,** 32-46.

ANDERSON, M. J., ELLINGSEN, K. E. & MCARDLE, B. H. 2006. Multivariate dispersion as a measure of beta diversity. *Ecology Letters,* 9**,** 683-693.

AUBENEAU, A., HANRAHAN, B., BOLSTER, D. & TANK, J. L. 2014. Substrate size and heterogeneity control anomalous transport in small streams. *Geophysical Research Letters,* 41**,** 8335-8341.

ÁVILA-ARCOS, M. C., CAPPELLINI, E., ROMERO-NAVARRO, J. A., WALES, N., MORENO-MAYAR, J. V., RASMUSSEN, M., FORDYCE, S. L., MONTIEL, R., VIELLE-CALZADA, J.-P., WILLERSLEV, E. & GILBERT, M. T. P. 2011. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Scientific Reports,* 1**,** 74.

BACHY, C., DOLAN, J. R., LÓPEZ-GARCÍA, P., DESCHAMPS, P. & MOREIRA, D. 2013. Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. *The ISME Journal,* 7**,** 244-255.

BAIRD, I. & FLAHERTY, M. 2005. Mekong River Fish Conservation Zones in Southern Laos: Assessing Effectiveness Using Local Ecological Knowledge. *Environmental Management,* 36**,** 439-454.

BAIRD, I. G. 2007. Local ecological knowledge and small-scale freshwater fisheries management in the Mekong River in southern Laos. *Fishers' Knowledge in Fisheries Science and Management. Haggan, N., Neis, B. and Baird, IG (eds). UNESCO, Paris***,** 247-266.

BAO, T. Q. 2001. *Local knowledge in the study of river fish biology : experiences from the Mekong,* Phnom Penh, Mekong River Commission.

BARAN, E., GUERIN, E. & SARAY, S. 2012. Features of high fish productivity in the Mekong Basin: case of 32 biodiversity hotspots. Report for the project "A Climate Resilient Mekong: maintaining the flows that nourish life" led by the Natural Heritage Institute. . Phnom Penh, Cambodia: WorldFish Center.

BARNES, M. A., TURNER, C. R., JERDE, C. L., RENSHAW, M. A., CHADDERTON, W. L. & LODGE, D. M. 2014. Environmental Conditions Influence eDNA Persistence in Aquatic Systems. *Environmental Science & Technology,* 48**,** 1819-1827.

BARNES, M. & TURNER, C. 2015. The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics,* 17**,** 1-17.

BARNOSKY, A. D., MATZKE, N., TOMIYA, S., WOGAN, G. O. U., SWARTZ, B., QUENTAL, T. B., MARSHALL, C., MCGUIRE, J. L., LINDSEY, E. L., MAGUIRE, K. C., MERSEY, B. & FERRER, E. A. 2011. Has the Earth's sixth mass extinction already arrived? *Nature,* 471**,** 51-57.

BAYLEY, P. B. & PETERSON, J. T. 2001. An Approach to Estimate Probability of Presence and Richness of Fish Species. *Transactions of the American Fisheries Society,* 130**,** 620-633.

BEJA-PEREIRA, A., OLIVEIRA, R., ALVES, P. C., SCHWARTZ, M. K. & LUIKART, G. 2009. Advancing ecological understandings through technological transformations in noninvasive genetics. *Mol Ecol Resour,* 9**,** 1279-301.

BELLEMAIN, PATRICIO, H. C., GRAY, T., GUEGAN, F., VALENTINI, A., MIAUD, C. & DEJEAN, T. 2016. Trails of river monsters: Detecting critically endangered Mekong giant catfish Pangasianodon gigas using environmental DNA. *Global Ecology and Conservation,* 7**,** 148-156.

BELLEMAIN, E., CARLSEN, T., BROCHMANN, C., COISSAC, E., TABERLET, P. & KAUSERUD, H. 2010. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC microbiology,* 10**,** 189.

BENSON, D. A., CAVANAUGH, M., CLARK, K., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL, J. & SAYERS, E. W. 2012. GenBank. *Nucleic acids research,* 41**,** D36-D42.

BIGGS, J., EWALD, N., VALENTINI, A., GABORIAUD, C., DEJEAN, T., GRIFFITHS, R. A., FOSTER, J., WILKINSON, J. W., ARNELL, A., BROTHERTON, P., WILLIAMS, P. & DUNN, F. 2015. Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (Triturus cristatus). *Biological Conservation,* 183**,** 19-28.

BISTA, I., CARVALHO, G. R., WALSH, K., SEYMOUR, M., HAJIBABAEI, M., LALLIAS, D., CHRISTMAS, M. & CREER, S. 2017. Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications,* 8**,** 14087.

BONAR, S. A., HUBERT, W. A. & WILLIS, D. W. 2009. Standard methods for sampling North American freshwater fishes.

BOWLER, J. M., JOHNSTON, H., OLLEY, J. M., PRESCOTT, J. R., ROBERTS, R. G., SHAWCROSS, W. & SPOONER, N. A. 2003. New ages for human occupation and climatic change at Lake Mungo, Australia. *Nature,* 421**,** 837-840.

BRUCE, K. D., HIORNS, W. D., HOBMAN, J. L., OSBORN, A. M., STRIKE, P. & RITCHIE, D. A. 1992. Amplification of DNA from native populations of soil bacteria by using the polymerase chain reaction. *Applied and Environmental Microbiology,* 58**,** 3413-3416.

BRYANT, J. A., STEWART, F. J., EPPLEY, J. M. & DELONG, E. F. 2012. Microbial community phylogenetic and trait diversity declines with depth in a marine oxygen minimum zone. *Ecology,* 93**,** 1659-1673.

BYLEMANS, J., FURLAN, E. M., HARDY, C. M., MCGUFFIE, P., LINTERMANS, M. & GLEESON, D. M. 2017. An environmental DNA-based method for monitoring spawning activity: a case study, using the endangered Macquarie perch (Macquaria australasica). *Methods in Ecology and Evolution,* 8**,** 646-655.

CAMPBELL, I. 2009. Chapter 1 - Introduction. *In:* BY, E. & CAMPBELL, I. C. (eds.) *The Mekong.* San Diego: Academic Press.

CAPELIN, M., KOHN, P. & HOFFENBERG, P. 1998. Land use, land cover and land degradation in the catchment of Moreton Bay. *Moreton Bay and catchment***,** 55-66.

CARIANI, A., MESSINETTI, S., FERRARI, A., ARCULEO, M., BONELLO, J. J., BONNICI, L., CANNAS, R., CARBONARA, P., CAU, A. & CHARILAOU, C. 2017. Improving the Conservation of Mediterranean Chondrichthyans: The ELASMOMED DNA Barcode Reference Library. *PloS one,* 12**,** e0170244.

CARIM, K. J., PADGETT-STEWART, T., WILCOX, T. M., YOUNG, M. K., MCKELVEY, K. S. & SCHWARTZ, M. K. 2015. Protocol for collecting environmental DNA samples from streams. *In:* SERVICE, U. F. (ed.). National Genomics Center for Wildlife and Fish Conservation.

CHAIN, F. J. J., BROWN, E. A., MACISAAC, H. J. & CRISTESCU, M. E. 2016. Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. *Diversity and Distributions,* 22**,** 493-504.

CHANG, C. H., SHAO, K. T., LIN, H. Y., CHIU, Y. C., LEE, M. Y., LIU, S. H. & LIN, P. L. 2017. DNA barcodes of the native ray-finned fishes in Taiwan. *Molecular ecology resources,* 17**,** 796-805.

CIVADE, R., DEJEAN, T., VALENTINI, A., ROSET, N., RAYMOND, J.-C., BONIN, A., TABERLET, P. & PONT, D. 2016. Spatial Representativeness of Environmental DNA Metabarcoding Signal for Fish Biodiversity Assessment in a Natural Freshwater System. *PLoS ONE,* 11**,** e0157366.

CLARKE, K. & WARWICK, R. 2001. Ordination of samples by multi-dimensional scaling (MDS). *Change in marine communities: An approach to statistical analysis and interpretation. Plymouth: Primer-E Ltd*.

CLARKE, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. *Austral Ecology,* 18**,** 117-143.

CLARKE, K. R. & GORLEY, R. N. 2006. *PRIMER V6: user manual-tutorial*, Plymouth Marine Laboratory.

COATES-MARNANE, J., OLLEY, J., BURTON, J. & SHARMA, A. 2016. Catchment clearing accelerates the infilling of a shallow subtropical bay in east coast Australia. *Estuarine, Coastal and Shelf Science,* 174**,** 27-40.

COLQUHOUN, D. 2014. An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society Open Science,* 1**,** 140216-140216.

CUNJAK, R. A., ROUSSEL, J. M., GRAY, M. A., DIETRICH, J. P., CARTWRIGHT, D. F., MUNKITTRICK, K. R. & JARDINE, T. D. 2005. Using stable isotope analysis with telemetry or mark-recapture data to identify fish movement and foraging. *Oecologia,* 144**,** 636-646.

DARLING, J. A. & MAHON, A. R. 2011. From molecules to management: adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environ Res,* 111**,** 978-88.

DE GRAAF, G. J., GRAINGER, R. J. R., WESTLUND, L., WILLMANN, R., MILLS, D., KELLEHER, K. & KORANTENG, K. 2011. The status of routine fishery data collection in Southeast Asia, central America, the South Pacific, and West Africa, with special reference to small-scale fisheries. *ICES Journal of Marine Science: Journal du Conseil,* 68**,** 1743-1750.

DEAGLE, B. E., EVESON, J. P. & JARMAN, S. N. 2006. Quantification of damage in DNA recovered from highly degraded samples – a case study on DNA in faeces. *Frontiers in Zoology,* 3**,** 11-11.

DEAGLE, B. E., THOMAS, A. C., SHAFFER, A. K., TRITES, A. W. & JARMAN, S. N. 2013. Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count? *Molecular Ecology Resources,* 13**,** 620-633.

DEAGLE, B. E., THOMAS, A. C., MCINNES, J. C., CLARKE, L. J., VESTERINEN, E. J., CLARE, E. L., KARTZINEL, T. R. & EVESON, J. P. 2018. Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular Ecology,* 00**,** 1-16.

DEINER, K. & ALTERMATT, F. 2014. Transport Distance of Invertebrate Environmental DNA in a Natural River. *PLoS ONE,* 9**,** e88786.

DEINER, K., WALSER, J.-C., MÄCHLER, E. & ALTERMATT, F. 2015. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation,* 183**,** 53-63.

DEINER, K., FRONHOFER, E. A., MÄCHLER, E., WALSER, J.-C. & ALTERMATT, F. 2016. Environmental DNA reveals that rivers are conveyer belts of biodiversity information. *Nature Communications,* 7**,** 12544.

DEINER, K., BIK, H. M., MÄCHLER, E., SEYMOUR, M., LACOURSIÈRE-ROUSSEL, A., ALTERMATT, F., CREER, S., BISTA, I., LODGE, D. M., DE VERE, N., PFRENDER, M. E. & BERNATCHEZ, L. 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology***,** n/a-n/a.

DEJEAN, T., VALENTINI, A., DUPARC, A., PELLIER-CUIT, S., POMPANON, F., TABERLET, P. & MIAUD, C. 2011. Persistence of Environmental DNA in Freshwater Ecosystems. *PLoS ONE,* 6**,** e23398.

DEJEAN, T., VALENTINI, A., MIQUEL, C., TABERLET, P., BELLEMAIN, E. & MIAUD, C. 2012. Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog Lithobates catesbeianus. *Journal of Applied Ecology,* 49**,** 953-959.

DENNELL, R. & PETRAGLIA, M. D. 2012. The dispersal of Homo sapiens across southern Asia: how early, how often, how complex? *Quaternary Science Reviews,* 47**,** 15-22.

DENNISON, W. C. & ABAL, E. G. 1999. *Moreton Bay study: a scientific basis for the healthy waterways campaign*, South East Qld Regional Water Quality Management Strategy Team.

DÍAZ-FERGUSON, E. E. & MOYER, G. R. 2014. History, applications, methodological issues and perspectives for the use environmental DNA (eDNA) in marine and freshwater environments. *2014,* 62**,** 12.

DICKIE, I. A. 2010. Insidious effects of sequencing errors on perceived diversity in molecular surveys. *The New Phytologist,* 188**,** 916-918.

DOI, H., INUI, R., AKAMATSU, Y., KANNO, K., YAMANAKA, H., TAKAHARA, T. & MINAMOTO, T. 2017. Environmental DNA analysis for estimating the abundance and biomass of stream fish. *Freshwater Biology,* 62**,** 30-39.

DUDGEON, D. 2000. The Ecology of Tropical Asian Rivers and Streams in Relation to Biodiversity Conservation. *Annual Review of Ecology and Systematics,* 31**,** 239-263.

DUDGEON, D., ARTHINGTON, A. H., GESSNER, M. O., KAWABATA, Z. I., KNOWLER, D. J., LÉVÊQUE, C., NAIMAN, R. J., PRIEUR-RICHARD, A. H., SOTO, D. & STIASSNY, M. L. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews,* 81**,** 163-182.

DUDGEON, D. 2011. Asian river fishes in the Anthropocene: threats and conservation challenges in an era of rapid environmental change. *J Fish Biol,* 79**,** 1487-524.

EATON, A. D., CLESCERI, L. S., GREENBERG, A. E., FRANSON, M. A. H., AMERICAN PUBLIC HEALTH, A., AMERICAN WATER WORKS, A. & WATER ENVIRONMENT, F. 1998. *Standard methods for the examination of water and wastewater,* Washington, DC, American Public Health Association.

EDGAR, R. C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Meth,* 10**,** 996-998.

EDGAR, R. C. 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv***,** 081257.

EICHMILLER, J. J., BAJER, P. G. & SORENSEN, P. W. 2014. The Relationship between the Distribution of Common Carp and Their Environmental DNA in a Small Lake. *PLoS ONE,* 9**,** e112611.

EICHMILLER, J. J., BEST, S. E. & SORENSEN, P. W. 2016a. Effects of Temperature and Trophic State on Degradation of Environmental DNA in Lake Water. *Environmental Science & Technology,* 50**,** 1859-1867.

EICHMILLER, J. J., MILLER, L. M. & SORENSEN, P. W. 2016b. Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Molecular Ecology Resources,* 16**,** 56-68.

ELBRECHT, V. & LEESE, F. 2015. Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass-Sequence Relationships with an Innovative Metabarcoding Protocol. *PLoS ONE,* 10**,** e0130324.

EVANS, N. T., OLDS, B. P., RENSHAW, M. A., TURNER, C. R., LI, Y., JERDE, C. L., MAHON, A. R., PFRENDER, M. E., LAMBERTI, G. A. & LODGE, D. M. 2016. Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources,* 16**,** 29-41.

EVANS, N. T., LI, Y., RENSHAW, M. A., OLDS, B. P., DEINER, K., TURNER, C. R., JERDE, C. L., LODGE, D. M., LAMBERTI, G. A. & PFRENDER, M. E. 2017a. Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences,* 74**,** 1362-1374.

EVANS, N. T., SHIREY, P. D., WIERINGA, J. G., MAHON, A. R. & LAMBERTI, G. A. 2017b. Comparative Cost and Effort of Fish Distribution Detection via Environmental DNA Analysis and Electrofishing. *Fisheries,* 42**,** 90-99.

EVANS, N. T. & LAMBERTI, G. A. 2018. Freshwater fisheries assessment using environmental DNA: A primer on the method, its potential, and shortcomings as a conservation tool. *Fisheries Research,* 197**,** 60-66.

FICETOLA, G. F., MIAUD, C., POMPANON, F. & TABERLET, P. 2008. Species detection using environmental DNA from water samples. *Biology Letters,* 4**,** 423-425.

FICETOLA, G. F., COISSAC, E., ZUNDEL, S., RIAZ, T., SHEHZAD, W., BESSIÈRE, J., TABERLET, P. & POMPANON, F. 2010. An in silico approach for the evaluation of DNA barcodes. *BMC genomics,* 11**,** 434-443.

FICETOLA, PANSU, J., BONIN, A., COISSAC, E., GIGUET-COVEX, C., DE BARBA, M., GIELLY, L., LOPES, C. M., BOYER, F., POMPANON, F., RAYÉ, G. & TABERLET, P. 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources,* 15**,** 543-556.

FILY, M. & D'AUBENTON, F. 1965. Cambodia. report on fisheries technology in the Great Lake and the Tonle Sap, 1962-63. Paris, France: National Museum of Natural History.

FISHBASE. 2016. *Search for family "Pangasiidae" on January 23, 2016.* [Online]. Available: http://www.fishbase.org/Nomenclature/NominalSpeciesList.php?Family=Pangasiidae.

FLATHER, C. H. & SIEG, C. H. 2007. Species rarity: definition, causes and classification. *In:* M.G. RAPHAEL, R. M. (ed.) *Conservation of rare or little-known species: Biological, social, and economic considerations.* Island Press.

FLYNN, J. M., BROWN, E. A., CHAIN, F. J. J., MACISAAC, H. J. & CRISTESCU, M. E. 2015. Toward accurate molecular identification of species in complex environmental samples: testing the performance of sequence filtering and clustering methods. *Ecology and Evolution,* 5**,** 2252-2266.

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. *Mol Mar Biol Biotechnol,* 3**,** 294-9.

FUKUMOTO, S., USHIMARU, A. & MINAMOTO, T. 2015. A basin-scale application of environmental DNA assessment for rare endemic species and closely related exotic species in rivers: a case study of giant salamanders in Japan. *Journal of Applied Ecology,* 52**,** 358-365.

FURLAN, E. M., GLEESON, D., HARDY, C. M. & DUNCAN, R. P. 2016. A framework for estimating the sensitivity of eDNA surveys. *Molecular Ecology Resources,* 16**,** 641-654.

GILES, F. 1935. An account of the ceremonies and rites performed when catchingthe pla buk: a species of catfish inhabiting the waters of the River Mekong, the Northern and Eastern frontier of Siam. *Natural History Bulletin of the Siam Society,* 28**,** 91-113.

GOLDBERG, C. S., PILLIOD, D. S., ARKLE, R. S. & WAITS, L. P. 2011. Molecular Detection of Vertebrates in Stream Water: A Demonstration Using Rocky Mountain Tailed Frogs and Idaho Giant Salamanders. *Plos One,* 6**,** e22746.

GOLDBERG, C. S., STRICKLER, K. M. & PILLIOD, D. S. 2015. Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. *Biological Conservation,* 183**,** 1-3.

GOLDBERG, C. S., TURNER, C. R., DEINER, K., KLYMUS, K. E., THOMSEN, P. F., MURPHY, M. A., SPEAR, S. F., MCKEE, A., OYLER-MCCANCE, S. J., CORNMAN, R. S., LARAMIE, M. B., MAHON, A. R., LANCE, R. F., PILLIOD, D. S., STRICKLER, K. M., WAITS, L. P., FREMIER, A. K., TAKAHARA, T., HERDER, J. E. & TABERLET, P. 2016. Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution,* 7**,** 1299-1307.

GRIFFITHS, E. & GUPTA, R. S. 2007. Identification of signature proteins that are distinctive of the Deinococcus-Thermus phylum. *Int Microbiol,* 10**,** 201-8.

GUNDERSEN, D. E. & LEE, I. M. 1996. Ultrasensitive detection of phytoplasmas by nested-PCR assays using two universal primer pairs. *Phytopathologia Mediterranea,* 35**,** 144-151.

HAGELBERG, E., THOMAS, M. G., COOK, C. E., JR., SHER, A. V., BARYSHNIKOV, G. F. & LISTER, A. M. 1994. DNA from ancient mammoth bones. *Nature,* 370**,** 333-4.

HAILE, J., HOLDAWAY, R., OLIVER, K., BUNCE, M., GILBERT, M. T. P., NIELSEN, R., MUNCH, K., HO, S. Y. W., SHAPIRO, B. & WILLERSLEV, E. 2007. Ancient DNA Chronology within Sediment Deposits: Are Paleobiological Reconstructions Possible and Is DNA Leaching a Factor? *Molecular Biology and Evolution,* 24**,** 982-989.

HAILE, J., FROESE, D. G., MACPHEE, R. D. E., ROBERTS, R. G., ARNOLD, L. J., REYES, A. V., RASMUSSEN, M., NIELSEN, R., BROOK, B. W., ROBINSON, S., DEMURO, M., GILBERT, M. T. P., MUNCH, K., AUSTIN, J. J., COOPER, A., BARNES, I., MÖLLER, P. & WILLERSLEV, E. 2009. Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences,* 106**,** 22352-22357.

HAJIBABAEI, M., SMITH, M. A., JANZEN, D. H., RODRIGUEZ, J. J., WHITFIELD, J. B. & HEBERT, P. D. N. 2006. A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes,* 6**,** 959-964.

HALLS, A. S., CONLAN, I. A., WISESJINDAWAT, K., PHOUTHAVONGS, S., VIRAVONG, S., CHAN, S. & VU, V. A. 2013a. Atlas of deep pools in the Lower Mekong River and some of its tributaries. Mekong River Commission, Phnom Penh, Cambodia

HALLS, A. S., PAXTON, B. R., HALL, N. & AL., E. 2013b. Integrated analysis of data from MRC Fisheries Monitoring Programmes in the Lower Mekong Basin.

HÄNFLING, B., LAWSON HANDLEY, L., READ, D. S., HAHN, C., LI, J., NICHOLS, P., BLACKMAN, R. C., OLIVER, A. & WINFIELD, I. J. 2016. Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular ecology,* 25**,** 3101-3119.

HECK, K. L., VAN BELLE, G. & SIMBERLOFF, D. 1975. Explicit Calculation of the Rarefaction Diversity Measurement and the Determination of Sufficient Sample Size. *Ecology,* 56**,** 1459-1461.

HERMOSO, V., KENNARD, M. J. & LINKE, S. 2013. Data Acquisition for Conservation Assessments: Is the Effort Worth It? *PLoS ONE,* 8**,** e59662.

HERMOSO, V., KENNARD, M. J. & LINKE, S. 2014. Assessing the risks and opportunities of presence-only data for conservation planning. *Journal of Biogeography,* 42**,** 218-228.

HERMOSO, V., LINKE, S., JANUCHOWSKI-HARTLEY, S. R. & KENNARD, M. J. 2015. *Freshwater conservation planning*, Cambridge University Press.

HIGUCHI, R., BOWMAN, B., FREIBERGER, M., RYDER, O. A. & WILSON, A. C. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature,* 312**,** 282-4.

HOFREITER, M., SERRE, D., POINAR, H. N., KUCH, M. & PAABO, S. 2001. Ancient DNA. *Nature Reviews Genetics,* 2**,** 353+.

HOFREITER, M., PAIJMANS, J. L. A., GOODCHILD, H., SPELLER, C. F., BARLOW, A., FORTES, G. G., THOMAS, J. A., LUDWIG, A. & COLLINS, M. J. 2015. The future of ancient DNA: Technical advances and conceptual shifts. *BioEssays,* 37**,** 284-293.

HOGAN, Z., MOYLE, P., MAY, B., VANDER ZANDEN, J. & BAIRD, I. 2004. The imperiled giants of the Mekong: ecologists struggle to understand-and protect-Southeast Asia's large, migratory catfish. *American Scientist,* 92**,** 228-237.

HOGAN, Z. 2011 Imperiled Giant Fish and Mainstream Dams in the Lower Mekong Basin: Assessment of Current Status, Threats, and Mitigation. University of Nevada, Reno. USA

HOGAN, Z. 2013a. *Catlocarpio siamensis* [Online]. The IUCN Red List of Threatened Species

HOGAN, Z. 2013b. *Pangasianodon gigas* [Online]. The IUCN Red List of Threatened Species. Available: http://www.iucnredlist.org/details/15944/0 [Accessed 10 November 2014].

HUGHES, A. O., OLLEY, J. M., CROKE, J. C. & MCKERGOW, L. A. 2009. Sediment source changes over the last 250 years in a dry-tropical catchment, central Queensland, Australia. *Geomorphology,* 104**,** 262-275.

HUNTER, M. E., DORAZIO, R. M., BUTTERFIELD, J. S. S., MEIGS-FRIEND, G., NICO, L. G. & FERRANTE, J. A. 2017. Detection limits of quantitative and digital PCR assays and their influence in presence–absence surveys of environmental DNA. *Molecular Ecology Resources,* 17**,** 221-229.

ILLUMINA 2016. Nextera DNA Library Prep: Protocol Guide

IUCN. 2018. *The IUCN Red List of Threatened Species.* [Online]. Available: <http://www.iucnredlist.org> [Accessed 2 February 2018].

JANE, S. F., WILCOX, T. M., MCKELVEY, K. S., YOUNG, M. K., SCHWARTZ, M. K., LOWE, W. H., LETCHER, B. H. & WHITELEY, A. R. 2015. Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Molecular Ecology Resources,* 15**,** 216-227.

JARMAN, S. N., MCINNES, J. C., FAUX, C., POLANOWSKI, A. M., MARTHICK, J., DEAGLE, B. E., SOUTHWELL, C. & EMMERSON, L. 2013. Adélie Penguin Population Diet Monitoring by Analysis of Food DNA in Scats. *PLOS ONE,* 8**,** e82227.

JENKINS, A., KULLANDER, F. F. & TAN, H. H. 2009. *The IUCN Red List of Threatened Species* [Online]. Available: <http://www.iucnredlist.org> [Accessed 2 February 2018].

JERDE, C. L., MAHON, A. R., CHADDERTON, W. L. & LODGE, D. M. 2011. "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters,* 4**,** 150-157.

JERDE, C. L., CHADDERTON, W. L., MAHON, A. R., RENSHAW, M. A., CORUSH, J., BUDNY, M. L., MYSOREKAR, S. & LODGE, D. M. 2013. Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. *Canadian Journal of Fisheries and Aquatic Sciences,* 70**,** 522-526.

JERDE, C. L., OLDS, B. P., SHOGREN, A. J., ANDRUSZKIEWICZ, E. A., MAHON, A. R., BOLSTER, D. & TANK, J. L. 2016. The influence of stream bottom substrate on the retention and transport of vertebrate environmental DNA. *Environmental Science & Technology,* 50**,** 8770-8779.

JI, Y., ASHTON, L., PEDLEY, S. M., EDWARDS, D. P., TANG, Y., NAKAMURA, A., KITCHING, R., DOLMAN, P. M., WOODCOCK, P., EDWARDS, F. A., LARSEN, T. H., HSU, W. W., BENEDICK, S., HAMER, K. C., WILCOVE, D. S., BRUCE, C., WANG, X., LEVI, T., LOTT, M., EMERSON, B. C. & YU, D. W. 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters,* 16**,** 1245-1257.

JONES, M. R. & GOOD, J. M. 2016. Targeted capture in evolutionary and ecological genomics. *Molecular Ecology,* 25**,** 185-202.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S., DURAN, C., THIERER, T., ASHTON, B., MEINTJES, P. & DRUMMOND, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-9.

KELLY, R. P., PORT, J. A., YAMAHARA, K. M. & CROWDER, L. B. 2014. Using Environmental DNA to Census Marine Fishes in a Large Mesocosm. *PLoS ONE,* 9**,** e86175.

KENNARD, M. J., PUSEY, B. J., HARCH, B. D., DORE, E. & ARTHINGTON, A. H. 2006. Estimating local stream fish assemblage attributes: sampling effort and efficiency at two spatial scales. *Marine and Freshwater Research,* 57**,** 635-653.

KÉRY, M. & SCHMIDT, B. 2008. Imperfect detection and its consequences for monitoring for conservation. *Community Ecology,* 9**,** 207-216.

KLYMUS, K. E., RICHTER, C. A., CHAPMAN, D. C. & PAUKERT, C. 2015. Quantification of eDNA shedding rates from invasive bighead carp Hypophthalmichthys nobilis and silver carp Hypophthalmichthys molitrix. *Biological Conservation,* 183**,** 77-84.

LACOURSIÈRE-ROUSSEL, A., CÔTÉ, G., LECLERC, V. & BERNATCHEZ, L. 2016a. Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. *Journal of Applied Ecology,* 53**,** 1148-1157.

LACOURSIÈRE-ROUSSEL, A., ROSABAL, M. & BERNATCHEZ, L. 2016b. Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. *Molecular Ecology Resources,* 16**,** 1401-1414.

LAMBERT, D. M., RITCHIE, P. A., MILLAR, C. D., HOLLAND, B., DRUMMOND, A. J. & BARONI, C. 2002. Rates of Evolution in Ancient DNA from Adélie Penguins. *Science,* 295**,** 2270-2273.

LAPOINTE, N. W. R., CORKUM, L. D. & MANDRAK, N. E. 2006. A Comparison of Methods for Sampling Fish Diversity in Shallow Offshore Waters of Large Rivers. *North American Journal of Fisheries Management,* 26**,** 503-513.

LARAMIE, M. B., PILLIOD, D. S. & GOLDBERG, C. S. 2015. Characterizing the distribution of an endangered salmonid using environmental DNA analysis. *Biological Conservation,* 183**,** 29-37.

LERAY, M., AND N. KNOWLTON 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *PNAS,* 112**,** 2076-2081.

LESLIE, C. & HANCOCK, G. 2008. Estimating the date corresponding to the horizon of the first detection of 137 Cs and 239+ 240 Pu in sediment cores. *Journal of environmental radioactivity,* 99**,** 483-490.

LODGE, D. M., TURNER, C. R., JERDE, C. L., BARNES, M. A., CHADDERTON, L., EGAN, S. P., FEDER, J. L., MAHON, A. R. & PFRENDER, M. E. 2012. Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology,* 21**,** 2555-2558.

LOH, J., GREEN, R. E., RICKETTS, T., LAMOREUX, J., JENKINS, M., KAPOS, V. & RANDERS, J. 2005. *The Living Planet Index: using species population time series to track trends in biodiversity*.
LOISELLE, B. A., HOWELL, C. A., GRAHAM, C. H., GOERCK, J. M., BROOKS, T., SMITH, K. G. & WILLIAMS, P. H. 2003. Avoiding Pitfalls of Using Species Distribution Models in Conservation Planning. *Conservation Biology,* 17**,** 1591-1600.
LORENZEN, K., SUKUMASAVIN, N. & HOGAN, Z. 2006. Development of a conservation strategy for the critically endangered Mekong giant catfish: quantitative assessment report. *Mekong Giant Catfish Working Group Report.* 3 ed.
MACHLER, E., DEINER, K., STEINMANN, P. & ALTERMATT, F. 2014. Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. *Freshwater Science,* 33**,** 1174-1183.
MÄCHLER, E., DEINER, K., SPAHN, F. & ALTERMATT, F. 2016. Fishing in the Water: Effect of Sampled Water Volume on Environmental DNA-Based Detection of Macroinvertebrates. *Environmental Science & Technology,* 50**,** 305-312.
MACKENZIE, D. I., NICHOLS, J. D., LACHMAN, G. B., DROEGE, S., ANDREW ROYLE, J. & LANGTIMM, C. A. 2002. ESTIMATING SITE OCCUPANCY RATES WHEN DETECTION PROBABILITIES ARE LESS THAN ONE. *Ecology,* 83**,** 2248-2255.
MAHON, A. R., BARNES, M. A., LI, F., EGAN, S. P., TANNER, C. E., RUGGIERO, S. T., FEDER, J. L. & LODGE, D. M. 2013a. DNA-based species detection capabilities using laser transmission spectroscopy. *Journal of the Royal Society Interface,* 10**,** 20120637.
MAHON, A. R., JERDE, C. L., GALASKA, M., BERGNER, J. L., CHADDERTON, W. L., LODGE, D. M., HUNTER, M. E. & NICO, L. G. 2013b. Validation of eDNA Surveillance Sensitivity for Detection of Asian Carps in Controlled and Field Experiments. *Plos One,* 8.
MAMANOVA, L., COFFEY, A. J., SCOTT, C. E., KOZAREWA, I., TURNER, E. H., KUMAR, A., HOWARD, E., SHENDURE, J. & TURNER, D. J. 2010. Target-enrichment strategies for next-generation sequencing. *Nat Meth,* 7**,** 111-118.
MARMION, M., PARVIAINEN, M., LUOTO, M., HEIKKINEN, R. K. & THUILLER, W. 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions,* 15**,** 59-69.
MARTELLINI, A., PAYMENT, P. & VILLEMUR, R. 2005. Use of eukaryotic mitochondrial DNA to differentiate human, bovine, porcine and ovine sources in fecally contaminated surface water. *Water Res,* 39**,** 541-8.
MARUYAMA, A., NAKAMURA, K., YAMANAKA, H., KONDOH, M. & MINAMOTO, T. 2014. The Release Rate of Environmental DNA from Juvenile and Adult Fish. *PLoS ONE,* 9**,** e114639.
MASON, V. C., LI, G., HELGEN, K. M. & MURPHY, W. J. 2011. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Research,* 21**,** 1695-1704.
MATISOO-SMITH, E., ROBERTS, K., WELIKALA, N., TANNOCK, G., CHESTER, P., FEEK, D. & FLENLEY, J. 2008. Recovery of DNA and pollen from New Zealand lake sediments. *Quaternary International,* 184**,** 139-149.
MATSUI, K., HONJO, M. & KAWABATA, Z. I. 2001. Estimation of the fate of dissolved DNA in thermally stratified lake water from the stability of exogenous plasmid DNA. *Aquatic Microbial Ecology,* 26**,** 95-102.
MAXWELL, W. 1970. The sedimentary framework of Moreton Bay, Queensland. *Marine and Freshwater Research,* 21**,** 71-88.
MINAMOTO, T., YAMANAKA, H., TAKAHARA, T., HONJO, M. & KAWABATA, Z. I. 2012. Surveillance of fish species composition using environmental DNA. *Limnology,* 13**,** 193-197.
MIYA, M., SATO, Y., FUKUNAGA, T., SADO, T., POULSEN, J. Y., SATO, K., MINAMOTO, T., YAMAMOTO, S., YAMANAKA, H., ARAKI, H., KONDOH, M. & IWASAKI, W. 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.
MOYER, G. R., DIAZ-FERGUSON, E., HILL, J. E. & SHEA, C. 2014. Assessing environmental DNA detection in controlled lentic systems. *PLoS One,* 9**,** e103767.

MRC 1996. Hydrographic Atlas of the Lower-Mekong River (1:20 000), Volumes I-V.

NATICK 2017. MATLAB R2017a ed.: The MathWorks

NCBI. *GenBank* [Online]. NCBI. Available: http://www.ncbi.nlm.nih.gov/genbank/ 2014].

NEEBLING, T. E. & QUIST, M. C. 2011. Comparison of Boat Electrofishing, Trawling, and Seining for Sampling Fish Assemblages in Iowa's Nonwadeable Rivers. *North American Journal of Fisheries Management,* 31**,** 390-402.

NEIL, D. T., ORPIN, A. R., RIDD, P. V. & YU, B. 2002. Sediment yield and impacts from river catchments to the Great Barrier Reef lagoon: a review. *Marine and Freshwater Research,* 53**,** 733-752.

NEWBOLD, J. D., THOMAS, S. A., MINSHALL, G. W., CUSHING, C. E. & GEORGIAN, T. 2005. Deposition, benthic residence, and resuspension of fine organic particles in a mountain stream. *Limnology and Oceanography,* 50**,** 1571-1580.

NGAMSIRI, T., NAKAJIMA, M., SUKMANOMON, S., SUKUMASAVIN, N., KAMONRAT, W., NA-NAKORN, U. & TANIGUCHI, N. 2007. Genetic diversity of wild mekong giant catfish Pangasianodon gigas collected from Thailand and Cambodia. *Fisheries Science,* 73**,** 792-799.

O'BRIEN, K., TUAZON, D., GRINHAM, A. & CALLAGHAN, A. 2012. *Impact of mud deposited by 2011 flood on marine and estuarine habitats in Moreton Bay*, Healthy Waterways Brisbane.

OGRAM, A., SAYLER, G. S. & BARKAY, T. 1987. The extraction and purification of microbial DNA from sediments. *Journal of Microbiological Methods,* 7**,** 57-66.

OKSANEN, J., BLANCHET, F. G., KINDT, R., LEGENDRE, P., MINCHIN, P. R., O'HARA, R., SIMPSON, G. L., SOLYMOS, P., STEVENS, M. H. H. & WAGNER, H. 2013. Package 'vegan'. *Community ecology package, version,* 2.

OLDEN, J. D. & JACKSON, D. A. 2002. A comparison of statistical approaches for modelling fish species distributions. *Freshwater Biology,* 47**,** 1976-1995.

OLDS, B. P., JERDE, C. L., RENSHAW, M. A., LI, Y., EVANS, N. T., TURNER, C. R., DEINER, K., MAHON, A. R., BRUESEKE, M. A. & SHIREY, P. D. 2016. Estimating species richness using environmental DNA. *Ecology and evolution,* 6**,** 4214-4226.

OLIVEIRA, L., KNEBELSBERGER, T., LANDI, M., SOARES, P., RAUPACH, M. & COSTA, F. 2016. Assembling and auditing a comprehensive DNA barcode reference library for European marine fishes. *Journal of fish biology,* 89**,** 2741-2754.

OLLEY, J. M., PIETSCH, T. & ROBERTS, R. G. 2004. Optical dating of Holocene sediments from a variety of geomorphic settings using single grains of quartz. *Geomorphology,* 60**,** 337-358.

OLSON, Z. H., BRIGGLER, J. T. & WILLIAMS, R. N. 2012. An eDNA approach to detect eastern hellbenders (Cryptobranchus a. alleganiensis) using samples of water. *Wildlife Research,* 39**,** 629-636.

PÄÄBO, S., HIGUCHI, R. G. & WILSON, A. C. 1989. Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *Journal of Biological Chemistry,* 264**,** 9709-12.

PÄÄBO, S., POINAR, H., SERRE, D., JAENICKE-DESPRES, V. & ET AL. 2004. GENETIC ANALYSES FROM ANCIENT DNA. *Annual Review of Genetics,* 38**,** 645-79.

PAWLOWSKI, J., ESLING, P., LEJZEROWICZ, F., CEDHAGEN, T. & WILDING, T. A. 2014. Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. *Molecular Ecology Resources,* 14**,** 1129-1140.

PEDERSEN, GINOLHAC, A., ORLANDO, L., OLSEN, J., ANDERSEN, K., HOLM, J., FUNDER, S., WILLERSLEV, E. & KJÆR, K. H. 2013. A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quaternary Science Reviews,* 75**,** 161-168.

PEDERSEN, OVERBALLE-PETERSEN, S., ERMINI, L., SARKISSIAN, C. D., HAILE, J., HELLSTROM, M., SPENS, J., THOMSEN, P. F., BOHMANN, K., CAPPELLINI, E., SCHNELL, I. B., WALES, N. A., CAROE, C., CAMPOS, P. F., SCHMIDT, A. M. Z., GILBERT, M. T. P., HANSEN, A. J., ORLANDO, L. & WILLERSLEV, E. 2015. Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society Biological Sciences,* 370**,** 20130383.

PEDERSEN, RUTER, A., SCHWEGER, C., FRIEBE, H., STAFF, R. A., KJELDSEN, K. K., MENDOZA, M. L. Z., BEAUDOIN, A. B., ZUTTER, C., LARSEN, N. K., POTTER, B. A., NIELSEN, R., RAINVILLE, R. A., ORLANDO, L., MELTZER, D. J., KJR, K. H. & WILLERSLEV, E. 2016. Postglacial viability and colonization in North Americas ice-free corridor. *Nature,* 537**,** 45.

PEREZ, C. R., BONAR, S. A., AMBERG, J. J., LADELL, B., REES, C., STEWART, W. T., GILL, C. J., CANTRELL, C. & ROBINSON, A. T. 2017. Comparison of American Fisheries Society (AFS) Standard Fish Sampling Techniques and Environmental DNA for Characterizing Fish Communities in a Large Reservoir. *North American Journal of Fisheries Management,* 37**,** 1010-1027.

PIETRAMELLARA, G., ASCHER, J., CECCHERINI, M. T., NANNIPIERI, P. & WENDEROTH, D. 2007. Adsorption of pure and dirty bacterial DNA on clay minerals and their transformation frequency. *Biology and Fertility of Soils,* 43**,** 731-739.

PILLIOD, D. S., GOLDBERG, C. S., ARKLE, R. S. & WAITS, L. P. 2013. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences,* 70**,** 1123-1130.

PILLIOD, D. S., GOLDBERG, C. S., ARKLE, R. S. & WAITS, L. P. 2014. Factors influencing detection of eDNA from a stream-dwelling amphibian. *Molecular Ecology Resources,* 14**,** 109-116.

POINAR, H. N., HOFREITER, M., SPAULDING, W. G., MARTIN, P. S., STANKIEWICZ, B. A., BLAND, H., EVERSHED, R. P., POSSNERT, G. & PÄÄBO, S. 1998. Molecular Coproscopy: Dung and Diet of the Extinct Ground Sloth (Nothrotheriops shastensis). *Science,* 281**,** 402-406.

POLZ, M. F. & CAVANAUGH, C. M. 1998. Bias in Template-to-Product Ratios in Multitemplate PCR. *Applied and Environmental Microbiology,* 64**,** 3724-3730.

POULSEN, A., POEU, O., VIRAVONG, S., SUNTORNRATANA, U. & TUNG, N. T. 2002. Deep pools as dry season fish habitats in the Mekong River basin. *MRC Technical Paper***,** 24.

POWELL, J. 1998. *People and trees : a thematic history of South East Queensland with particular reference to forested areas 1823-1997 / Judith Powell,* Barton, ACT ; Indooroopilly, Qld, Forests Taskforce, Department of the Prime Minister and Cabinet ; Regional Forest Assessments, Dept. of Natural Resources, Qld.

PUSEY, B., KENNARD, M. & ARTHINGTON, A. 2004. *Freshwater Fishes of North-Eastern Australia,* Collingwood, VIC CSIRO PUBLISHING.

R CORE TEAM 2016. R: A language and environment for statistical computing *R Foundation for Statistical Computing.* Vienna, Austria

RATNASINGHAM, S. & HEBERT, P. D. N. 2007. bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes,* 7**,** 355-364.

RAWLENCE, N. J., LOWE, D. J., WOOD, J. R., YOUNG, J. M., CHURCHMAN, G. J., HUANG, Y. T. & COOPER, A. 2014. Using palaeoenvironmental DNA to reconstruct past environments: progress and prospects. *Journal of Quaternary Science,* 29**,** 610-626.

REES, H. C., MADDISON, B. C., MIDDLEDITCH, D. J., PATMORE, J. R. M. & GOUGH, K. C. 2014. REVIEW The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology,* 51**,** 1450-1459.

REVENGA, C., CAMPBELL, I., ABELL, R., DE VILLIERS, P. & BRYER, M. 2005. Prospects for monitoring freshwater ecosystems towards the 2010 targets. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 360**,** 397-413.

RICCIARDI, A. & RASMUSSEN, J. B. 1999. Extinction Rates of North American Freshwater Fauna. *Conservation Biology,* 13**,** 1220-1222.

RITCHIE, P. A., MILLAR, C. D., GIBB, G. C., BARONI, C. & LAMBERT, D. M. 2004. Ancient DNA Enables Timing of the Pleistocene Origin and Holocene Expansion of Two Adélie Penguin Lineages in Antarctica. *Molecular Biology and Evolution,* 21**,** 240-248.

ROBSON, H. L. A., NOBLE, T. H., SAUNDERS, R. J., ROBSON, S. K. A., BURROWS, D. W. & JERRY, D. R. 2016. Fine-tuning for the tropics: application of eDNA technology for invasive fish detection in tropical freshwater ecosystems. *Molecular Ecology Resources,* 16**,** 922-932.

RONI, P., HANSON, K. & BEECHIE, T. 2008. Global Review of the Physical and Biological Effectiveness of Stream Habitat Rehabilitation Techniques. *North American Journal of Fisheries Management,* 28**,** 856-890.

ROSE, P. M., KENNARD, M. J., MOFFATT, D. B., SHELDON, F. & BUTLER, G. L. 2016. Testing Three Species Distribution Modelling Strategies to Define Fish Assemblage Reference Conditions for Stream Bioassessment and Related Applications. *PLoS ONE,* 11**,** e0146728.

ROUSSEL, J.-M., PAILLISSON, J.-M., TRÉGUIER, A. & PETIT, E. 2015. The downside of eDNA as a survey tool in water bodies. *Journal of Applied Ecology,* 52**,** 823-826.

SAGAN, L. 1967. On the origin of mitosing cells. *Journal of theoretical biology,* 14**,** 225-274.

SAMBROOK, J. & RUSSELL, D. W. 2006. Purification of Nucleic Acids by Extraction with Phenol:Chloroform. *Cold Spring Harbor Protocols,* 2006**,** pdb.prot4455.

SASSOUBRE, L. M., YAMAHARA, K. M., GARDNER, L. D., BLOCK, B. A. & BOEHM, A. B. 2016. Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine Fish. *Environ Sci Technol,* 50**,** 10456-10464.

SCHELLER, R. M., SNARSKI, V. M., EATON, J. G. & OEHLERT, G. W. 1999. An Analysis of the Influence of Annual Thermal Variables on the Occurrence of Fifteen Warmwater Fishes. *Transactions of the American Fisheries Society,* 128**,** 257-264.

SCHMELZLE, M. C. & KINZIGER, A. P. 2016. Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. *Molecular Ecology Resources,* 16**,** 895-908.

SCHMIDT, B. R., KERY, M., URSENBACHER, S., HYMAN, O. J. & COLLINS, J. P. 2013. Site occupancymodels in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. *Methods in Ecology and Evolution,* 4**,** 646-653.

SCHNELL, I. B., BOHMANN, K. & GILBERT, M. T. P. 2015. Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources,* 15**,** 1289-1303.

SCHULTZ, M. T. & LANCE, R. F. 2015. Modeling the Sensitivity of Field Surveys for Detection of Environmental DNA (eDNA). *PLoS ONE,* 10**,** e0141503.

SEFC, K. M., PAYNE, R. B. & SORENSON, M. D. 2007. Single base errors in PCR products from avian museum specimens and their effect on estimates of historical genetic diversity. *Conservation Genetics,* 8**,** 879-884.

SHAW, J. L. A., CLARKE, L. J., WEDDERBURN, S. D., BARNES, T. C., WEYRICH, L. S. & COOPER, A. 2016. Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biological Conservation,* 197**,** 131-138.

SHOGREN, A. J., TANK, J. L., ANDRUSZKIEWICZ, E., OLDS, B., MAHON, A. R., JERDE, C. L. & BOLSTER, D. 2017. Controls on eDNA movement in streams: Transport, Retention, and Resuspension. *Scientific Reports,* 7**,** 5065.

SIGSGAARD, E. E., CARL, H., MOLLER, P. R. & THOMSEN, P. F. 2015. Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. *Biological Conservation,* 183**,** 46-52.

SILVANO, R. A. M., SILVA, A. L., CERONI, M. & BEGOSSI, A. 2008. Contributions of ethnobiology to the conservation of tropical rivers and streams. *Aquatic Conservation: Marine and Freshwater Ecosystems,* 18**,** 241-260.

SIMMONS, M., TUCKER, A., CHADDERTON, W. L., JERDE, C. L. & MAHON, A. R. 2015. Active and passive environmental DNA surveillance of aquatic invasive species. *Canadian Journal of Fisheries and Aquatic Sciences,* 73**,** 76-83.

SLON, V., HOPFE, C., WEIß, C. L., MAFESSONI, F., DE LA RASILLA, M., LALUEZA-FOX, C., ROSAS, A., SORESSI, M., KNUL, M. V., MILLER, R., STEWART, J. R., DEREVIANKO, A. P., JACOBS, Z., LI, B., ROBERTS, R. G., SHUNKOV, M. V., DE LUMLEY, H., PERRENOUD, C., GUŠIĆ, I., KUĆAN, Ž., RUDAN, P., AXIMU-PETRI, A., ESSEL, E., NAGEL, S., NICKEL, B., SCHMIDT, A., PRÜFER, K., KELSO, J., BURBANO, H. A., PÄÄBO, S. & MEYER, M. 2017. Neandertal and Denisovan DNA from Pleistocene sediments. *Science,* 356**,** 605-608.

SMITH, D. P. & PEAY, K. G. 2014. Sequence Depth, Not PCR Replication, Improves Ecological Inference from Next Generation DNA Sequencing. *PLOS ONE,* 9**,** e90234.

SODHI, N., POSA, M., LEE, T., BICKFORD, D., KOH, L. & BROOK, B. 2010. The state and conservation of Southeast Asian biodiversity. *Biodiversity and Conservation,* 19**,** 317-328.

SOMMERIA-KLEIN, G., ZINGER, L., TABERLET, P., COISSAC, E. & CHAVE, J. 2016. Inferring neutral biodiversity parameters using environmental DNA data sets. *Scientific Reports,* 6**,** 35644.

STARNBACH, M. N., FALKOW, S. & TOMPKINS, L. S. 1989. Species-specific detection of Legionella pneumophila in water by DNA amplification and hybridization. *Journal of Clinical Microbiology,* 27**,** 1257-1261.

STEIN, E. D., MARTINEZ, M. C., STILES, S., MILLER, P. E. & ZAKHAROV, E. V. 2014. Is DNA Barcoding Actually Cheaper and Faster than Traditional Morphological Methods: Results from a Survey of Freshwater Bioassessment Efforts in the United States? *PLOS ONE,* 9**,** e95525.

STRAYER, D. L. & DUDGEON, D. 2010. Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society,* 29**,** 344-358.

STRICKLER, K. M., FREMIER, A. K. & GOLDBERG, C. S. 2015. Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation,* 183**,** 85-92.

TABERLET, GRIFFIN, S., GOOSSENS, B., QUESTIAU, S., MANCEAU, V., ESCARAVAGE, N., WAITS, L. P. & BOUVET, J. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research,* 24**,** 3189-3194.

TABERLET, P., COISSAC, E., HAJIBABAEI, M. & RIESEBERG, L. H. 2012a. Environmental DNA. *Molecular Ecology,* 21**,** 1789-1793.

TABERLET, P., COISSAC, E., POMPANON, F., BROCHMANN, C. & WILLERSLEV, E. 2012b. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology,* 21**,** 2045-2050.

TAKAHARA, T., MINAMOTO, T., YAMANAKA, H., DOI, H. & KAWABATA, Z. I. 2012. Estimation of Fish Biomass Using Environmental DNA. *PLoS ONE,* 7**,** e35868.

TAKAHARA, T., MINAMOTO, T. & DOI, H. 2013. Using Environmental DNA to Estimate the Distribution of an Invasive Fish Species in Ponds. *PLoS One,* 8**,** e56584.

TANG, C. Q., LEASI, F., OBERTEGGER, U., KIENEKE, A., BARRACLOUGH, T. G. & FONTANETO, D. 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 16208-16212.

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, D. J., LEE, S. S., LARSSON, K.-H., KOHOUT, P., HOSAKA, K., HIIESALU, I., HENKEL, T. W., HAREND, H., GUO, L.-D., GRESLEBIN, A., GRELET, G., GEML, J., GATES, G., DUNSTAN, W., DUNK, C., DRENKHAN, R., DEARNALEY, J., DE KESEL, A., 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.

THOMSEN, P. F., KIELGAST, J., IVERSEN, L. L., MØLLER, P. R., RASMUSSEN, M. & WILLERSLEV, E. 2012a. Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples. *PLoS ONE,* 7**,** e41732.

THOMSEN, P. F., KIELGAST, J. O. S., IVERSEN, L. L., WIUF, C., RASMUSSEN, M., GILBERT, M. T. P., ORLANDO, L. & WILLERSLEV, E. 2012b. Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology,* 21**,** 2565-2573.

THOMSEN, P. F. & WILLERSLEV, E. 2015. Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation,* 183**,** 4-18.

THORNE, A., GRÜN, R., MORTIMER, G., SPOONER, N. A., SIMPSON, J. J., MCCULLOCH, M., TAYLOR, L. & CURNOE, D. 1999. Australia's oldest human remains: age of the Lake Mungo 3 skeleton. *Journal of Human Evolution,* 36**,** 591-612.

THUILLER, W., LAFOURCADE, B., ENGLER, R. & ARAÚJO, M. B. 2009. BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography,* 32**,** 369-373.

TURNER, C. R., UY, K. L. & EVERHART, R. C. 2015. Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation,* 183**,** 93-102.

ULIBARRI, R. M., BONAR, S. A., REES, C., AMBERG, J., LADELL, B. & JACKSON, C. 2017. Comparing Efficiency of American Fisheries Society Standard Snorkeling Techniques to Environmental DNA Sampling Techniques. *North American Journal of Fisheries Management,* 37**,** 644-651.
VALENTINI, A., POMPANON, F. & TABERLET, P. 2009. DNA barcoding for ecologists. *Trends in Ecology & Evolution,* 24**,** 110-117.
VALENTINI, A., TABERLET, P., MIAUD, C., CIVADE, R., HERDER, J., THOMSEN, P. F., BELLEMAIN, E., BESNARD, A., COISSAC, E., BOYER, F., GABORIAUD, C., JEAN, P., POULET, N., ROSET, N., COPP, G. H., GENIEZ, P., PONT, D., ARGILLIER, C., BAUDOIN, J. M., PEROUX, T., CRIVELLI, A. J., OLIVIER, A., ACQUEBERGE, M., LE BRUN, M., MOLLER, P. R., WILLERSLEV, E. & DEJEAN, T. 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol Ecol,* 25**,** 929-42.
VALIERE, N. & TABERLET, P. 2000. Urine collected in the field as a source of DNA for species and individual identification. *Molecular Ecology,* 9**,** 2150-2152.
VIRAVONG, S., PHOUNSAVATH, S., PHOTITAY, C., PUTREA, S., CHAN, S., KOLDING, J., VALBO-JØRGENSEN, J. & PHOUTAVONG, K. 2006. Hydroacoustic surveys of deep pools in Southern Lao PDR and Northern Cambodia. *MRC Technical Report Series***,** 76.
VÖRÖSMARTY, C. J., MCINTYRE, P. B., GESSNER, M. O., DUDGEON, D., PRUSEVICH, A., GREEN, P., GLIDDEN, S., BUNN, S. E., SULLIVAN, C. A. & REIDY LIERMANN, C. 2010. Global threats to human water security and river biodiversity. *nature,* 467**,** 555.
WAITS, L. P. & PAETKAU, D. 2005. NONINVASIVE GENETIC SAMPLING TOOLS FOR WILDLIFE BIOLOGISTS: A REVIEW OF APPLICATIONS AND RECOMMENDATIONS FOR ACCURATE DATA COLLECTION. *Journal of Wildlife Management,* 69**,** 1419-1433.
WANG, Q., GARRITY, G. M., TIEDJE, J. M. & COLE, J. R. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology,* 73**,** 5261-5267.
WEAVER, M. J., MAGNUSON, J. J. & CLAYTON, M. K. 1993. Analyses for Differentiating Littoral Fish Assemblages with Catch Data from Multiple Sampling Gears. *Transactions of the American Fisheries Society,* 122**,** 1111-1119.
WILCOX, T. M., MCKELVEY, K. S., YOUNG, M. K., JANE, S. F., LOWE, W. H., WHITELEY, A. R. & SCHWARTZ, M. K. 2013. Robust Detection of Rare Species Using Environmental DNA: The Importance of Primer Specificity. *PLoS ONE,* 8**,** e59520.
WILCOX, T. M., MCKELVEY, K. S., YOUNG, M. K., SEPULVEDA, A. J., SHEPARD, B. B., JANE, S. F., WHITELEY, A. R., LOWE, W. H. & SCHWARTZ, M. K. 2016. Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char Salvelinus fontinalis. *Biological Conservation,* 194**,** 209-216.
WILLERSLEV, E., HANSEN, A. J., CHRISTENSEN, B., STEFFENSEN, J. P. & ARCTANDER, P. 1999. Diversity of Holocene life forms in fossil glacier ice. *Proceedings of the National Academy of Sciences,* 96**,** 8017-8021.
WILLERSLEV, E., HANSEN, A. J., BINLADEN, J., BRAND, T. B. & ET AL. 2003. Diverse plant and animal genetic records from holocene and pleistocene sediments. *Science,* 300**,** 791-5.
WILLERSLEV, E., HANSEN, A. J. & POINAR, H. N. 2004. Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology & Evolution,* 19**,** 141-147.
WILLOUGHBY, J. R., WIJAYAWARDENA, B. K., SUNDARAM, M., SWIHART, R. K. & DEWOODY, J. A. 2016. The importance of including imperfect detection models in eDNA experimental design. *Molecular Ecology Resources,* 16**,** 837-844.
WILSON, K. A., WESTPHAL, M. I., POSSINGHAM, H. P. & ELITH, J. 2005. Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biological Conservation,* 122**,** 99-112.
WINEMILLER, K. O., MCINTYRE, P. B., CASTELLO, L., FLUET-CHOUINARD, E., GIARRIZZO, T., NAM, S., BAIRD, I. G., DARWALL, W., LUJAN, N. K., HARRISON, I., STIASSNY, M. L. J., SILVANO, R. A. M., FITZGERALD, D. B., PELICICE, F. M., AGOSTINHO, A. A., GOMES, L. C., ALBERT, J. S., BARAN, E., PETRERE, M., ZARFL, C., MULLIGAN, M., SULLIVAN, J. P., ARANTES, C. C., SOUSA, L. M., KONING, A. A., HOEINGHAUS, D. J., SABAJ, M., LUNDBERG, J. G., ARMBRUSTER, J., THIEME,

M. L., PETRY, P., ZUANON, J., VILARA, G. T., SNOEKS, J., OU, C., RAINBOTH, W., PAVANELLI, C. S., AKAMA, A., SOESBERGEN, A. V. & SÁENZ, L. 2016. Balancing hydropower and biodiversity in the Amazon, Congo, and Mekong. *Science,* 351**,** 128-129.

YE, J., COULOURIS, G., ZARETSKAYA, I., CUTCUTACHE, I., ROZEN, S. & MADDEN, T. L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics,* 13**,** 134.

YOCCOZ, N. G., NICHOLS, J. D. & BOULINIER, T. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution,* 16**,** 446-453.

YU, D. W., JI, Y., EMERSON, B. C., WANG, X., YE, C., YANG, C. & DING, Z. 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution,* 3**,** 613-623.

ZHAN, A., BAILEY, S. A., HEATH, D. D. & MACISAAC, H. J. 2014. Performance comparison of genetic markers for high-throughput sequencing-based biodiversity assessment in complex communities. *Molecular Ecology Resources,* 14**,** 1049-1059.

ZHANG, J., KOBERT, K., FLOURI, T. & STAMATAKIS, A. 2013. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics,* 30**,** 614-620.

ZON, G. & GEISER, T. G. 1991. Phosphorothioate oligonucleotides: chemistry, purification, analysis, scale-up and future directions. *Anticancer Drug Des,* 6**,** 539-68.