DOCTOR OF PHILOSOPHY

Exploring the biodiversity of the lakes of the Malay Archipelago using environmental DNA metabarcoding

Evans, Alice

Award date:
2019

Awarding institution:
Bangor University
University of Copenhagen

Exploring the biodiversity of the lakes of the Malay Archipelago using environmental DNA metabarcoding

A thesis submitted for the degree of Doctor of Philosophy,

School of Biological Sciences, Bangor University &

Faculty of Science, University of Copenhagen

Double PhD degree

Alice Ruth Evans

2019

Summary

The freshwater ecosystems of Southeast Asia are some of the most highly threatened in the world, due to anthropogenic impact from climate change, deforestation, the creation of hydropower dams and over-harvesting. Rapid, cost-effective and reliable monitoring of biodiversity is essential for the conservation of the exceptional biotic richness within this region. The emerging field of environmental DNA (eDNA) monitoring, using trace cells or fragments of DNA released into an environment to assign species to locations has potential to provide this type of information.

In this thesis, I explore the use of eDNA metabarcoding for monitoring freshwater aquatic biodiversity within Southeast Asia, focusing on fishes within the lakes of the Malay Archipelago. Firstly, I co-led to a published review of the field of eDNA in which we discuss how the field has developed, address current challenges, and predict future developments. Secondly, I conducted sampling of lakes across the Malaysian Peninsula as an initial exploration into the use of eDNA in tropical freshwaters using the ethanol precipitation method of environmental DNA collection, as well as conducted a mesocosm experiment to test eDNA degradation. Thirdly, after initial trouble shooting, I tested options for isolation and storage of aquatic eDNA to inform best practice solutions for eDNA field researchers, and found that the use of an enclosed filter system combined with a preservation buffer was the best approach. Fourthly, I conducted intensive sampling of a lake in Indonesia to investigate the dynamics of eDNA information within a tropical lentic environment, and found heterogenous detection of extant biodiversity. Finally, I undertook a large-scale biogeography study of the lakes of the Malay Archipelago, sampling from western Sumatra across to eastern Sulawesi using a filter approach for environmental DNA collection. Metabarcoding of aquatic eDNA samples was then employed for all samples, with a combination of primers targeting different mitochondrial regions to achieve a broad scope of biodiversity information. From the data, I recovered native, endemic and rare species, as well as introduced and invasive species linked to fisheries, aquaculture, the ornamental trade and pest-control. Overall, aquatic eDNA metabarcoding demonstrated great potential, allowing ecosystem level species detection, but further work on eDNA distribution, improvements to barcoding capabilities and the reliability of quantification, will greatly deepen the possibilities presented by aquatic eDNA metabarcoding in advancing wildlife and biodiversity monitoring in tropical habitats.

Danish Summary

Sydøstasiens ferskvandsystemer er nogle af de mest truede i verden på grund af menneskeskabte udfordringer fra klimaforandringer, skovrydning, oprettelse af dæmninger og udpining af jorde. Hurtig, effektiv og pålidelig overvågning af biodiversiteten er afgørende for bevarelsen af den unikke artsrigdom i denne region. Det stadig voksende forskningsfelt indenfor miljø-DNA (også kaldet eDNA), hvor man undersøger de fragmenter af DNA, der frigives af organismer til miljøet, har potentialet til at tilvejebringe denne ønskede artsinformation. I denne afhandling undersøger jeg brugen af DNA metabarcoding og miljø-DNA til overvågning af biodiversitet i ferskvand i Sydøstasien med fokus på fisk i søerne i det Malaysiske øhav. Først præsenterer jeg en offentliggjort gennemgang af området eDNA, hvor vi diskuterer, hvordan feltet har udviklet sig, løser aktuelle udfordringer og forudsiger den fremtidige udvikling. Yderligere gennemførte jeg prøveudtagninger fra søer fra den Malaysiske halvø for at undersøge brugen af miljø-DNA i tropisk ferskvand ved brug af en ethanol-udfældningsmetode for miljø-DNA-indsamlingen, samt udførte forsøg for at teste DNA-nedbrydningen af miljø-DNA. Selvom dette arbejde ikke gav pålidelige resultater, og derfor ikke er medtaget i denne afhandling, gav den stor erfaring i forhold til at implementere miljø-DNA-prøveudtagninger i troperne. Derefter testede jeg mulighederne for at isolere og opbevare akvatisk miljø-DNA for at finde frem til den bedste og mest praktiske løsning og fandt ud af, at brugen af et lukket filtersystem kombineret med en bevaringsbuffer var den bedste tilgang. Jeg gennemførte også en intensiv prøveudtagning af en sø i Indonesien for at undersøge dynamikken af miljø-DNA-information inden for et tropisk lentisk miljø og kunne påvise forekomsten af den eksisterende biodiversitet. Endelig gennemførte jeg en stor undersøgelse af søerne i det Malaysiske øhav, med stikprøver fra det vestlige Sumatra over til det østlige Sulawesi, ved brug af en filter-tilgang til miljø-DNA-indsamlingen. DNA metabarcoding af de akvatiske miljø-DNA-prøver blev derefter anvendt, med en kombination af primere rettet mod forskellige mitokondrieområder, for at fokusere på et bredt udvalg af biodiversiteten. Dette gav information om lokale, endemiske og sjældne arter samt introducerede og invasive arter knyttet til fiskeri, akvakultur, prydplanter og skadedyrsbekæmpelse. Samlet set viste DNA metabarcoding af det akvatiske miljø-DNA et stort potentiale til at påvise arter tilhørende forskellige økosystemer. Dog vil fremtidigt arbejde med miljø-DNA-fordelingen, forbedring af ”barcoding”-evnerne og pålideligheden af kvantificering fra miljø-DNA i høj grad kunne udvikle mulighederne yderligere i forhold til at bruge akvatisk miljø-DNA til overvågning af akvatisk biodiversitet i Sydøstasien.

Acknowledgements

As with all scientific advances, the work I've done has been built on the foundation laid previously by others, and guided by those around me who have lent me their expertise and advice. I have been extremely fortunate to work with labs across the world, with scientists from host of different nationalities, and have learned so much from their varying knowledge bases, competencies, skills and philosophies. This project was funded by the Natural Environment Research Council, with additional funding from the British Council and Chester Zoo.

Firstly, I would like to thank my supervisors, Dr Mark de Bruyn, Prof Si Creer, Prof Gary Carvalho, Prof Tom Gilbert, Prof Siti Azizah Mohd Nor and Prof Gusti Ngurah Mahardika for their invaluable contribution to my PhD.

Mark - thank you for sparking these ideas when I first contacted MEFGL wanting to create a PhD project, and guiding my academic and logistical directions. You always welcomed me popping into your office with a smile for chats, even when you were a new Dad and needing intravenous coffee after two hours' sleep. You listened to my ideas, panics and worries with patience and support both in the UK and from Australia, led me in publishing my first PhD paper, helped me find funding, run international training courses, and were a lot of fun to work with along the way ("More mojitos!") so thank you for all that you have done for me and this project which would not have happened without you.

Si – thank you for your consistent support and advice, particularly lending your expertise from the world of eDNA, experimental design, and data analysis. You have made a point of letting me know that your door is always open for me, and have given me the pushes I needed to keep focused. From sending me recipes, to debating the setup of metabarcoding plates, your input and guidance has been extremely appreciated.

Gary – thank you for your seasoned supervision throughout my PhD, in grounding me in the broader ecological science, in academic wisdom to formulate what is needed for my PhD, and especially in your compassionate support and advice. You are a privilege to work with, set an example for how academics can interact with humanity and passion, and have made my days at MEFGL a pleasure.

Tom – thank you for creating the opportunity for me to come to Denmark and work in your fantastic group, which allowed most of my PhD to be possible. I often joke that you must secretly have Hermione's time turner to have the ability to support the number of people in The Gilbert Group that you do, and fulfill all your scientific, academic, media-related,

family and extracurricular commitments. You are visionary academic and inspirational leader, teaching collaboration and kindness by example. Thank you for always making time for me, creating solutions at the click of your fingers, easing the logistics of migrating to Denmark, and bolstering me in the midst of gin-induced-Christmas-party-PhD-tears.

Siti – terima kasih banyak for hosting me at USM and facilitating my Malaysian field work. You made me feel like a member of the family whilst I was in Penang and made every possible effort to provide what I needed, from a place to rest my head, to a place to dig mesocosm holes in the ground, to a fantastic research team to support my sampling. You set a wonderful example to me as a leading female scientist, and it was an absolute pleasure to have the privilege of your company and advice. You are clearly loved by your brilliant group for good reason. I hope I will one day be able to return for a visit and have debates and laughter over roti canai in the market.

Ngurah – terma kasih banyak for accepting our request to partner in our venture through the Global Innovation Initiative and support me throughout my work in Indonesia, you allowed us to work in this beautiful and biodiverse region of the world. You were right when you said you would act as my Indonesian Dad and stopped me from killing myself on a scooter, as well as helping me find my place on Kertha Dalem, travelling with me to Bogor, imparting your philosophy on academia and life on our car rides, and advising me when we thought I had a tropical disease! Thank you also for your academic assistance in hosting our training course, allowing me my own mini 'eDNA lab', and arranging my field work support with your talented group.

Secondly, I would like to thank the other people who have given me essential academic and personal help in Bangor, Copenhagen, Malaysia and Indonesia.

From Bangor, thank you to Jess for being my PhD rock from before day one, being the best flat mate, acting as my human Filofax, making me laugh when I wanted to cry, and spurring me on when I was in 'The Valley of Shit'. From the early years, thank you to Alison, Sam, Tom, Caitlin, Hazel, Jonathan, Kitty, Will, Owain, Greg, Bryony, Antonia and Alix, for your friendship, support and company for much needed Friday pub beers and adventures in North Wales. To Nik for your statistical advice over a beer and constantly joyful presence, to Iliana for eDNA advice, to Jane for your organization wizardry, to Wendy for all-round-lab-magic, to Georgina for providing your R and statistical knowledge, and to Ian for bioinformatic tips. Thank you also to Georgina, James, Kitty, Jen and Will for hosting me in my final write up stage visits to Bangor, with delicious cooking and much needed

friendship. Also to Anita for acting as my internal for most of my PhD and a point of advice for postgrad issues and providing a platform to discuss the oft-forgotten issue of gender equality in academia, and Wolfgang for acting as my chair.

From Copenhagen, a huge thank you to Kristine for your writing skills on our TREE paper, metabarcoding advice, bioinformatics pipelines, teaching sessions on our training courses, and all-round-PhD-motivation including gems like 'forget the plankton!' and 'just decide, and DO IT' – I would have been completely lost without your guidance, and have been very lucky to work with your unstoppable force. To Martin for getting me started in the lab when I had forgotten how to hold a pipette, and sitting with me under the hood for hours whilst we went through protocols and judged what to do using my 'female intuition'. To Lis for your shining smile in the Danish darkness, giving me DAMe tips and coming to teach for us in Bangor, you are a force to be reckoned with! To Micaela for sharing your knowledge of Sterivex methodology, coming to teach in Bali and leading our MEE paper (also to Johan, Steen, Sarah, Mita, Peter, Ian and Eva and Philip for your essential contributions). Thank you to Felipe and Shyam for your bioinformatics help, to Philip, Eva and Ida for letting me take primer aliquots and giving me eDNA advice, to Steen for qPCR tips, and to all of the above as well as Lara, Aldo, Jazmin, Viktor, Kathi, Mikkel, Ana, George, Sama, Natalie, Haris and Inger, for the in and out of work fun times in Copenhagen.

From Malaysia, thank you to Din, Lia, Jam, Danial, Ana, Lim, Adiba, Wani and Nurul for your amazing help in getting my samples all across Malaysia, your endless positivity and work ethic was so inspiring. I realise now more than ever how much you went out of your way to help me, and I am so grateful. I had so much fun working in your group, eating our way around Penang, hanging on for dear life in the back of the van driving across the country, and taking shelter in the oasis of air con at the malls. Thank you for welcoming me with such warmth.

From Indonesia, thank you to my main accomplice, Andre, for your constant assistance and efficiency, and going above and beyond helping me with anything and everything I needed. Whether we were stuck in Semarang because of a volcano, making deals with the boatmen, doing extractions or eating until we couldn't move - you made my face hurt with laughing so much and were the best research partner I could ask for. To Yuli for getting additional samples for me and posting them to Denmark, as well as helping arrange our training course and anything I needed, and for being a wonderful friend. To the above, as well as Eka, Elok, Asa, Fajar, Afrita, Samsul, Fiah, Dhyan, Nana, Sutrisna, Jeanne and Dio for all your incredible help whilst I was in Bali, driving me around on the back of a scooter,

finding lab equipment, getting sore arms after hours of Sterivex filtering, and arranging our eDNA course. Thanks to Ibu Citra and Pak Anugra for helping me arrange sampling in Sulawesi and Kalimantan.

Thirdly, I would like to thank anyone outside of work who has supported me in my PhD ups and downs. To my Mum and Dad for supporting my decision to do a PhD, moving me back and forth across the country, and providing support when things got tough. To my sister, Bex, for the cuddles, dancing and dog videos which brightened the times between work, to Rose, Lucy S, Lucy F, Adam, Emma, Vicki, Tillique, Lis, Amy and Vikki for making the effort to visit me in Denmark and Bangor, and to all my other friends who put up with me being away, and were still there for me. Last but certainly not least, to Richard, my partner in crime, recently promoted to husband, for your endless love, support and advice, listening to my moaning, regularly flying across Europe to see me, bringing tea and cuddles when you know I'm stressed, sending my wonderful birthday video to Malaysia, flying to Indonesia to see me and being my unrelenting cheerleader – you have kept me going on this PhD journey, so the biggest thank you to you for consistently being my best friend.

PhD timeline and additional PhD activities

October 2013 – March 2014 – *began the PhD*: I completed compulsory courses and assignments for the postgrad module ONS2004 including a literature review, statistics course, project proposal, project presentation and scientific talk abstracts. I contributed to a review paper as joint-first author, published in *Trends in Ecology and Evolution* (Appendix 1; Bohmann *et al*. 2014), now at 358 citations. I wrote two grant proposals and received funding from both, consisting of £2,950 from the conservation research grant from Chester Zoo, and £147,995 from the Global Innovation Initiative (GII) from the British Council. I collected preliminary test samples in collaboration with staff from the Anglesey Sea Zoo and Chester Zoo to explore different filtering methods and ethanol precipitation methods on water from tanks with known species. I also attended PhD training courses run by Hutchinson Training & Development, on Project Management, Managing Your Supervisory Relationship, and Rapid Reading, and applied for a Malaysian Research Permit.

April 2014 – June 2014 – *first field work expedition*: I prepared equipment for eDNA sampling using the ethanol precipitation methodology and flew to Kuala Lumpur to finalise the research permit process. I created a mesocosm experiment using 45 L containers x 12 dug into the ground, inoculated with pond water, and populated with fish. I sampled the water before adding fish, every few days whilst fish were in the water, and for some time after removing the fish to test for eDNA degradation time. I collected eDNA samples from Lake Bukit Mera (04/05/2014, 18/05/2014, 11/02/2014), Tasik Pedu (20/05/2014), Tasik Bera (27/05/2014, 17/06/2014), Tasik Chini (28/05/2014, 18/06/2014) and Tasik Chenderoh (30/05/2014). I also performed extractions on all samples from the mesocosm experiment and lake sampling at the University of Sience Malaysia.

July 2014 – September 2014 – *training courses and preparation for moving to Copenhagen.* I attended a Climate Change and Species Distribution Modelling Course at Copenhagen University, August 2014, followed by the iEOS2014 International Environmental Omics Synthesis conference, Liverpool, September 2014 and the Environmental DNA Working Group Meeting, Hull University, September 2014 (where I presented a poster).

October 2014 – March 2015 – *first round of lab work in Copenhagen.* I learned the PCR optimisation, exploration of sample quality, metabarcoding PCR and library prep process using samples from Malaysia. I combined a small number of samples for sequencing with other people's sequencing runs to examine data, but sequences very poor quality and amplification generally unsuccessful. I applied for an Indonesian Research Permit for second

field work expedition. I formulated an experiment to compare eDNA sampling and storage methods, organised sampling in collaboration with others at Copenhagen University, and performed all extractions for this experiment. I also attended the NERC-MDIBL Environmental Genomics and Metabolomics, Birmingham, UK, March 2015.

April 2015 – May 2015 – *preparing for Indonesian field expedition*. I invited teachers, (Mark de Bruyn, Kristine Bohmann, Micaela Hellström, Pierre Taberlet, Eric Coissac) prepared the lecture schedule, invited participants and organised logistics for the eDNA training course in Bali. I prepared equipment for sampling, and documents for Indonesian Research Permit. I also presented a poster at the College of Natural Sciences PhD Conference, (and won best poster), May 2015.

June 2015 – August 2015 – *second field work expedition.* I ran an eDNA training course (an obligation of the GII grant) on the sampling, extraction and analysis of eDNA samples, and gave lectures during this course for participants from all across Indonesia. I sampled eDNA from lakes using the Sterivex filter method described in Chapter 2, with help from staff members of the Indonesian Biodiversity Research Unit (IBRC). I sampled Lake Tamblingan, Lake Beratan, Lake Buayn, Lake Batur, Lake Matano, Lake Semayang, Lake Melintang, and Lake Rawa Pening, whilst other IBRC members sampled additional lakes. I then performed extractions on all samples. I visited the Indonesian Government Research branch, LIPI, in Bogor to try to finalise research collaboration to fulfil permit requirements by presenting research to panel of scientists from the department of Limnology. I also attended the DIPnet workshop: Molecular Ecology and Biodiversity Informatics in Southeast Asia, Bali, Indonesia, August 2015.

September 2015 – December 2015 – *training course and preparation for Copenhagen:* I hosted a training course in Bangor on Environmental DNA and Bioinformatic Analysis for participants from the UK, USA and Indonesia (as part of the GII grant obligations). I organised and executed this training course, including inviting teachers (Mark de Bruyn, Si Creer, Kristine Bohmann, Lissandra Zepeda-Mendoza, Aurelie Bonin, Céline Mercier), organised the teaching schedule, delivered some lectures, organised sampling of sea water on a boat trip from Puffin Island, and taught eDNA extraction at the labs in MEFGL. I also formulated the idea and organised the UK's eDNA WGM to coincide with this course, by using the GII grant to fund this meeting, invite participants, organise meeting talk schedule, catering and entertainment, and assist in the running of the meeting of over 70 national and international academic and stakeholder delegates. After the training course and meeting, I focused on writing for a manuscript prepared by Mark de Bruyn on threats to freshwater

biodiversity in Southeast Asia and recommendations for their conservation (Appendix 5) submitted to *Bioscience*, and writing for the eDNA methods comparison experiment (Chapter 2). I also prepared to move to Copenhagen for the second round of lab work.

January 2016 – June 2016 – *second round of lab work in Copenhagen*: I performed PCR optimisation, exploration of eDNA samples, metabarcoding PCR, library prep and sequencing of all Indonesian and Malaysian samples, including previously sampled Malaysian lakes using ethanol precipitation. I rewrote the manuscript for *Bioscience* on threats to freshwater biodiversity in Southeast Asia based on reviewer's comments. I trained PhD student, Cátia Pereira from the University of Evora in eDNA sampling, and visited Evora, Portugal, to help her set up her eDNA experiment using mesocosm tanks. I began training in the DAMe pipeline to analyse eDNA metabarcoding data. I also presented a talk at Bangor University's College of Natural Sciences PhD Day and won best talk (May 2016).

August 2016 – December 2016 – *training, analysis and rerun of lab work*: I attended the International Society of Limnology Congress SIL2016, Torino, Italy, August 2016. I began data analysis using the DAMe pipeline. I returned to Copenhagen to redo library prep and sequencing for failed run of COI data, and do further training for bioinformatic analysis. I completed the week-long University Pedagogy Training Course, Copenhagen, October 2016. I presented a poster at the DNA Working Group Meeting, Edinburgh, UK, November 2016.

January 2017 – December 2017 – *Analysis, training and writing*: I worked through the metabarcoding analysis pipeline and focused on writing for thesis. I completed the training courses: Introduction to R, Statistical Modelling in R, Programming in R and Advanced Graphics in R, April 2017. I attended th International Barcode of Life conference (iBOL) (oral presentation), Kruger National Park, South Africa, November 2017. I attended the DNA Working Group Meeting, Salford, UK, December 2017. I assisted with a preliminary sampling trip to Loch Ness for an eDNA project spearheaded by Otago University with Neil Gemmell and Gert-Jan Jeunen in December 2017.

January 2018 – January 2019 – *Analysis and writing:* processed OTU tables, analysed read matches, performed statistical analysis and writing for the thesis.

Thesis outline and contributions

Chapter 1: General Introduction: environmental DNA for wildlife biology and biodiversity monitoring in Southeast Asia

Chapter 1 introduces the field of environmental DNA (eDNA), discusses how techniques and breadth of information has improved, and suggests the challenges faced within the field of eDNA research as well as solutions to overcome them. eDNA based methods are explored within the context of Southeast Asian freshwater ecosystems, focusing on three key areas: monitoring of invasive species, understanding ecosystem processes, and informing conservation management. The first half of this chapter is based on the review paper published in *Trends in Ecology & Evolution* in 2014, shown in Appendix 1 (Bohmann *et al*. 2014, Appendix 1) of which I am joint first author. Chapter 1 uses this paper as a starting point, updated to include research and developments published up to October 2018. The second half of this chapter, focusing on Southeast Asia, includes elements of a manuscript being prepared for submission to *Conservation Letters* of which I am first author (previously submitted to *Bioscience,* Appendix 3) which has received positive comments from the Editor as a pre-submission enquiry.

Chapter 2: Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter.

Chapter 2 explores the effect of different isolation techniques, storage techniques, and storage time on aquatic eDNA samples from a Danish lake based on qPCR amplification of two key fish species. This chapter was published in *Methods in Ecology and Evolution* in 2016 in a paper on which I am joint-first author (Appendix 2). For this study, we compared various eDNA filter materials and ethanol precipitation as potential capture methods, along with various preservation buffers and freezing as potential storage methods. I was part of the formulation of the idea for this study, planned the experimental design, led the sampling day, completed all extractions, and helped with the writing of the manuscript.

Chapter 3: Universal methods

This chapter describes the methods used in Chapter 4 and 5, as these chapters used the same sampling approach, molecular workflow and bioinformatic pipeline.

Chapter 4: The distribution of eDNA within the Indonesian lake, Danau Tamblingan: recommendations for eDNA sampling of tropical lentic habitats.

Chapter 4 tests for differences in taxonomic community composition generated from metabarcoding OTUs, and OTU richness between different sites of the same lake. When sampling eDNA from lacustrine habitats, it is unclear how many samples should be collected, and how far apart they should be collected to encompass the extant biodiversity. I sampled a small Balinese caldera lake at regular intervals across the surface and at different depth points, and used eDNA metabarcoding using a 12S, 16S and COI primer set, sequenced on the Illumina MiSeq. The fish and mammal species detected could be explained by previous studies. I found that taxonomic communities and OTU richness varied between points, and that this was affected by sample depth. However, further work testing points at more regular intervals, and storing filters in a buffer could increase the taxonomic information generated and give a clearer picture of how eDNA is spatially distributed within a tropical lake.

Chapter 5: Assessment of the aquatic biodiversity of the lakes of the Malay Archipelago using eDNA metabarcoding.

Chapter 5 explores the use of aquatic eDNA metabarcoding in assessing the extant biodiversity of a variety of lakes across the Malay Archipelago. Using a transect approach, subsamples were collected at regular transect intervals, combined into one large sample, and replicate filtrations performed from these combined subsamples to maximise the lake area covered. I recovered native, endemic, rare, introduced, invasive, ornamental and pest-control fish species; domestic, native and rare mammal species and a range of freshwater microfauna, meiofauna and microalgae which could be explained by the literature. OTU community composition and OTU richness was affected by altitude, lake area, maximum lake depth and trophic productivity. Further work on the lakes of the Malay Archipelago using a more intensive sampling approach across a larger area per lake, as well as adding more lakes, would help illuminate patterns influencing biodiversity such as anthropogenic impact.

Chapter 6: General discussion

Chapter 6 discusses the main findings of this thesis, what improvements could be made and places this work within a wider context.

Table of contents

List of figures

List of Tables

Glossary

Amplicon: a targeted fragment of DNA or RNA created by replication events or amplification, either naturally or artificially, through e.g. PCR.

Ancient DNA (aDNA): DNA extracted from specimens that have not been intentionally preserved for genetic analysis. Such samples are typically low quality and can include specimens from museum collections, archaeological finds, and subfossil remains of tissues or other DNA-containing sources (e.g., coprolites, hair).

Aphotic zone: the layer of a lake beneath the euphotic zone where light levels are too low for photosynthesizers, usually found within the hypolimnion or sometimes the metalimnion, consisting of light levels of less than 1% of the lake surface.

Barcode gap: the break between intra- and interspecific pairwise distances that underpins the success of DNA barcoding

Benthic zone: the zone between the lacustrine sediment and the water column, with a surface layer abundant with organisms.

Bioassessment/biomonitoring: the characterisation of ecosystem health using biological surveys through the detection of resident 'indicator' biota—including fish, insects, algae, plants and others.

Blocking primer: an oligonucleotide used to bind to DNA and overlap the primer-binding sites, so that amplification of the undesired species is prevented.

Bulk DNA: DNA obtained from community samples targeting particular organisms, such as from plankton collected with a plankton tow or large organisms scraped from rocks or collected in grabs.

Capture based aquaculture (CBA): wild caught juveniles such as milkfish *Chanos chanos* are reared in ponds, cages, and pens, which can be described as 'fisheries driven'.

Chimera: sequences that arise during amplification combining DNA fragments from two or more individuals.

Cloning: The process of producing genetically identical copies of an organism, either naturally or artificially. Cloning commonly refers to the insertion of DNA into a vector molecule (e.g. a plasmid) prior to selection for a gene of interest, DNA extraction and sequencing.

Community DNA: DNA derived from many individuals of different species.

Culture based fisheries (CBF): a form of aquaculture and conventional aquaculture, such as cage and pen culture. The CBF strategies involve stocking of hatchery-reared fish fingerlings into small natural and quasi-natural waterbodies. CBF can be described as 'aquaculture driven' in contrast to capture-based aquaculture, and is considered more environmentally friendly due to low addition of supplementary feeds.

Degenerate primers/universal primers: Primers used for amplicon sequencing where the targeted gene(s) is typically similar, but not identical.

Environmental DNA (eDNA): broadly speaking, eDNA is DNA sampled from an environment without first isolating the target organism. This may be in the form of intracellular or extracellular DNA from intraorganismal or extraorganismal sources. Some authors argue a stricter definition of 'true' eDNA, which is trace fragments or cells sampled from an abiotic environment without first isolating, or detecting signs of, the target organism.

Environmental RNA (eRNA): rather than deoxyribonucleaic acid targeted in eDNA samples, eRNA (environmental RNA - ribonucleic acid) deteriorates rapidly after cell death, likely providing a more accurate representation of viable communities.

Epilimnion: the upper, wind-mixed layer of a thermally stratified lake, turbulently mixed and exchanges gases with the atmosphere.

Euphotic zone: the layer of a lake directly beneath the surface usually found within the epilimnion, which supports photosynthesizers as light levels are ≥1% of the lake surface.

Eutrophic: trophic state of lakes with abundant nutrients e.g. phosphorous and nitrogen, high plant biomass (phytoplankton, algae, vascular plants) and undesirable water-quality characteristics (low transparency, green colour, odorous, low oxygen).

Exact amplicon sequence variants (ASVs): unique sequences as opposed to OTUs

Extended barcode: a species identification barcode based on an entire organelle genome and nuclear ribosomal DNAs.

Extracellular eDNA: eDNA located outside of the cell.

Extraorganismal eDNA: eDNA found outside of the target organism, i.e. eDNA in its most strict form, found as trace cells (intracellular eDNA) or trace fragments (extracellular eDNA).

Floating Net Cages (FNC): cages used to house fish for aquaculture, suspended at the surface of a lake or the ocean, known in Indonesia as keramba.

Genome skimming: the use of shallow-pass shotgun sequencing of genomic DNA to generate extended barcodes, simultaneously recovering all standard barcoding regions as well as other loci, and a link with all other phylogenetically informative genomic regions.

Genomic DNA: DNA extracted from an individual or collection of individuals of the same species.

Hypolimnion (plural noun: hypolimnia): the bottom, most dense layer of the lake, coldest in the summer and warmest in the winter, isolated from turbulent mixing and usually too dark for photosynthesis to occur.

Intracellular eDNA: eDNA located inside of the cell

Intraorganismal eDNA: eDNA found within the target organism, e.g. DNA of microbes within a soil sample, or DNA of nematodes within a benthic sediment sample.

Locus: The specific location of a gene or DNA sequence on a chromosome.

Invertebrate DNA (iDNA): invertebrate-derived DNA

Lacustrine: relating to or associated with lakes.

Limnetic (pelagic) zone: the inner open water portion of the lake away from the near shore area, where light does not usually penetrate to the bottom benthic zone, including the surface and bottom of the lake; the entire area of the lake after the littoral zone.

Littoral zone: near shore area where sunlight penetrates down to the sediment, with light levels of at least 1% of that at the lake surface, allowing growth of aquatic plants (macrophytes).

Marker gene: A gene or DNA sequence targeted in amplicon sequencing to screen for a specific organism group or functional gene.

Meromictic: describes a lake with layers that do not mix.

Mesotrophic: trophic state of lakes with medium level nutrients, with features in between eutrophic and oligotrophic states.

Metalimnion: the middle transitory layer of the lake, between the epilimnion and the hypolimnion, of medium density.

Metabarcoding: Use of gene-specific PCR primers to amplify DNA from a collection of organisms or from environmental DNA. Another term for amplicon sequencing.

Metagenetics (ecogenetics): the analysis of community taxon richness via the detection of homologous genes

Metagenomics (ecogenomics): sequencing of the total DNA extracted from a sample containing many different organisms. The random sequencing of gene fragments isolated from environmental samples, allowing sequencing of uncultivable organisms.

Metatranscriptomics: the study of metatranscriptomes, which comprise only expressed regions of the genomes present in eDNA samples. Shotgun sequencing of total RNA from

environmental samples. Techniques such as poly-A amplification or rRNA depletion are often used to target messenger (mRNA) transcripts to assess gene expression patterns in complex communities.

Microarray: a set of DNA sequences representing the entire set of genes of an organism, arranged in a grid pattern for use in genetic testing.

Microbiome: the microorganisms in a particular environment (e.g. the body or a part of the body).

Mitochondrial metagenomics (mito-metagenomics / MMG): a methodology for shotgun sequencing of total DNA from specimen mixtures and subsequent bioinformatic extraction of mitochondrial sequences.

Mitogenome: The sum of the genetic information contained in the chromosome of the mitochondrion.

Next generation sequencing (NGS)/high-throughput sequencing (HTS): the sequencing of many DNA fragments in parallel, using a number of different modern sequencing technologies including: Illumina (Solexa) sequencing, Ion torrent: Proton / PGM sequencing and SOLiD sequencing.

NTU: Nephelometric Turbidity Unit, used to measure turbidity through scattered light.

Oligotrophic: trophic state of lakes with low nutrients e.g. phosphorous and nitrogen, suppression of plant growth through scarce phosphorous, low dissolved carbon, high transparency, blue colour, oxygen retention, supporting fish and other eukaryotes.

Operational taxonomic unit (OTU): the taxonomic level of sampling defined by the researcher in a study; for example, individuals, populations, species, genera, or strains. OTUs are generated by comparing sequences to form a distance matrix, followed by clustering groups of sequences with a specified amount of variability allowed within each OTU.

PCR bias: – the differential PCR amplification of DNA fragments found in higher concentrations in the sample.

Polymerase chain reaction (PCR): Used to amplify a targeted piece of DNA, generating many copies of that particular DNA sequence.

Shotgun sequencing: DNA is fragmented into small segments which are individually sequenced and then reassembled into longer, continuous sequences using sequence assembly software.

Abbreviations

ADAS = Agricultural Development and Advisory Service

BLAST = Basic Local Alignment Search Tool

BOLD = Barcode of Life Data Systems

CBF = Culture Based Fisheries

COI = cytochrome *c* oxidase 1 mitochondrial gene

Dryad = Dryad Digital Repository

eDNA = Environmental DNA

iBOL = International Barcode of Life Project

NCBI = National Centre for Biotechnology Information

NGS = Next-generation sequencing

Chapter 1

General Introduction: Environmental DNA for wildlife biology and biodiversity monitoring in Southeast Asia

Chapter 1: General Introduction

1.1 Environmental DNA: the next generation of biodiversity monitoring

In 1966, the writers of Star Trek introduced intergalactic battles, alien invaders, and technology beyond the realm of reality. When the handheld Tricorder was used by Spock to test unexplored habitats, little did the writers know that the sci-fi technology to analyse an environment and its living components from a small sample would become a reality in just 50 Earth years. Free DNA molecules are ubiquitous, released from surface cells, internal fluids and waste material from plants and animals, and are collectively referred to as environmental DNA (eDNA). Any given environmental sample, whether water, air or soil, will contain a myriad of eDNA, and the information contained therein is now accessible owing to advances in sample preparation and NGS technology. Today, such perspectives of science fiction are a reality as a growing number of biologists are using eDNA for species detection and biomonitoring, circumventing, or at least alleviating, the need to sight or sample living organisms. Such approaches can accelerate the rate of discovery, as no *a priori* information about the likely species found in a particular environment is required to identify those species. Those working on invasive species, community and ecosystem processes underpinning biodiversity and functional diversity, and wildlife and conservation biology are likely to benefit the most from adoption of eDNA techniques.

1.2 A brief history of eDNA

The term 'environmental DNA' was first used in microbiology (Ogram *et al*. 1987) to explain the method of extracting DNA from an environmental sample of soil without first isolating the target microorganisms. This grew from analysing bacterial evolution (Woese *et al*. 1987), to revealing unknown microbial genetic diversity in extreme habitats (Pace *et al*. 1997), to shotgun sequencing whole genomes of aquatic marine microbial life (Venter *et al*. 2004), sparking a revolution of research on eukaryotic diversity, evolutionary relationships and ecology. As techniques became easier, cheaper and more widely known, eDNA methods were adopted in a range of fields, using a host of different techniques (Taberlet *et al*. 2012a). The growth of references which mention environmental DNA and metabarcoding (with their relation to fish in particular) is shown in Figure 1.1 below.

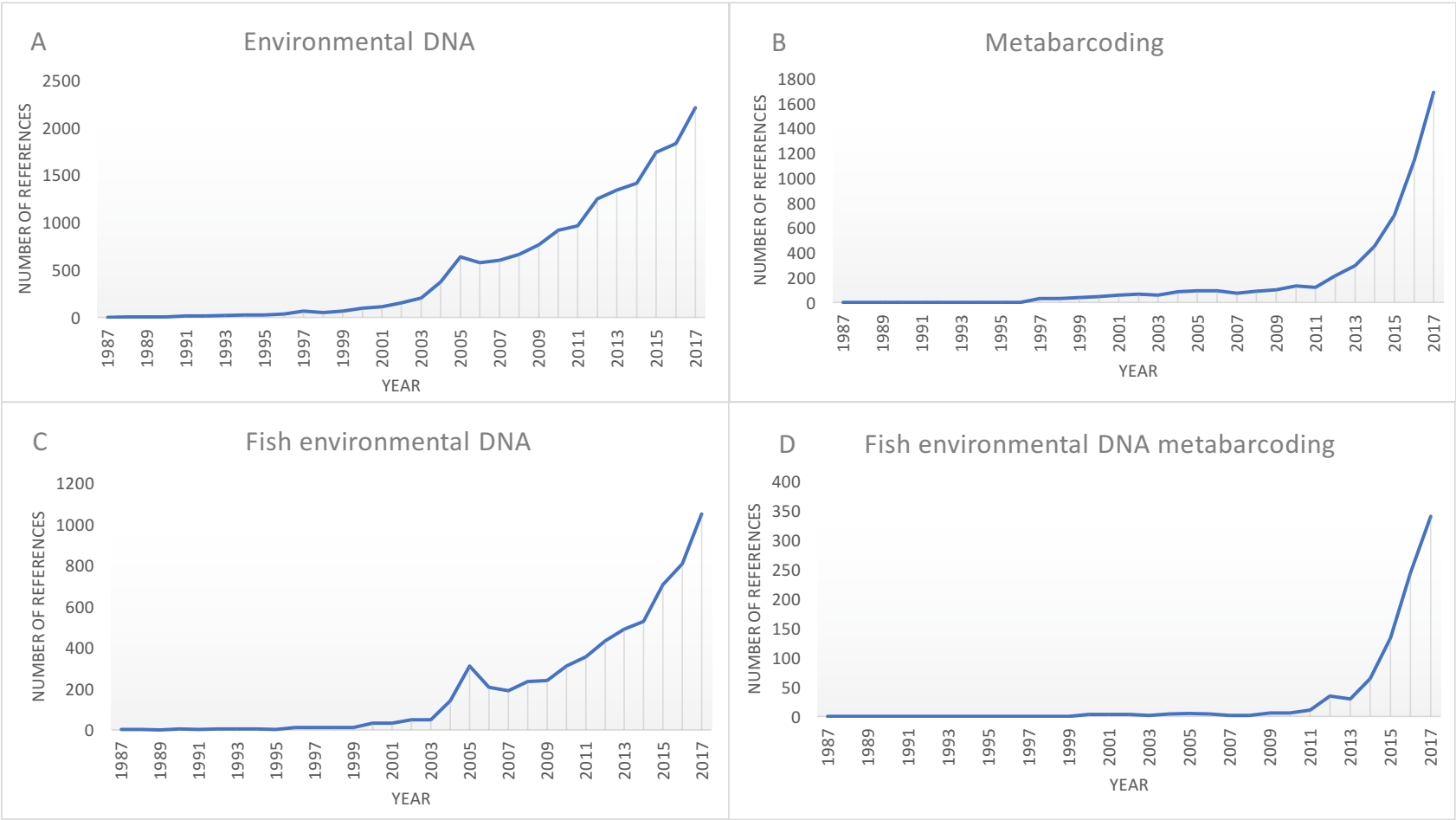


Figure 1.1. Environmental DNA research published from 1987 to 2017. Created using a Google Scholar search of publications with the exact words found in the graph titles searched using quotation marks and not including patents or citations. Each search counted results per year. A: "environmental DNA" B: "metabarcoding" OR "meta-barcoding" C: "fish" +and "environmental dna". D: "fish" + "environmental dna" + "metabarcoding" OR "meta-barcoding".

Accessing macrobial, rather than microbial, genetic information from environmental samples grew initially from the field of ancient DNA (aDNA) which used ancient ice, permafrost or sediment to detect animals and communities extinct for thousands of years (Willerslev *et al*. 2003). Studies focusing on species detection for dietary analysis from faecal samples have been performed for some time (e.g. Reed *et al*. 1997), although this type of sampling can be referred to as 'molecular scatology' rather than true eDNA (discussed below). Other such types of early eDNA samples included snow (Dalén *et al*. 2007), honey (Schnell *et al*. 2010) and browsed twigs (Nichols *et al*. 2012), but most eDNA sampling focused on soil or water. Contemporary eDNA sampling for macrobial life from water by Martellini *et al*. (2005), detected human, pig, cow and sheep mitochondrial DNA from river water running off farmland. Ficetola *et al*. (2008) then used eDNA to detect the invasive American Bullfrog (*Rana catesbeiana*) from pond water in France, which ignited a stream of aquatic eDNA studies for the detection of macrobial species. Since then, many eDNA and metabarcoding sample types have been collected for a range of different applications, organisms and habitats, highlighted in several reviews over the last five years (Lodge *et al*. 2012; Yoccoz, 2012a; Taberlet *et al*. 2012a; Taberlet *et al*. 2012b; Rees *et al*. 2014; Bohmann *et al*. 2014; Rees *et al*. 2015; Pedersen *et al*. 2015; Lawson Handley, 2015; Thomsen and Willerslev, 2016; Deiner *et al*. 2017b; Evans *et al*. 2017c; Hansen *et al*. 2018; Cristescu and Hebert, 2018).

1.3 What *is* eDNA?

Environmental DNA can most simply be defined as 'DNA obtained directly from environmental samples without first isolating the target organism, the predominant sources of which are from faeces, urine and epidermal cells, found free floating in an environment such as water, or persist, adsorbed in organic or inorganic particles (Dejean *et al*. 2011; Thomsen *et al*. 2012a). There is however a degree of ambiguity surrounding the definition of what environmental DNA is, and some debate focusing on what qualifies as true 'eDNA'. For example, Mahon *et al*. (2013) define eDNA as "dissolved DNA and/or fragments of tissue containing DNA". Based on this definition, it could be argued that DNA left behind on the tip of a feather, the surface of an egg shell, around faeces, or in a visibly large chunk of tissue (e.g. Amos *et al*. 1992) is environmental DNA, regardless of where or how it is found, as it does not involve trapping or catching the target species. On the other hand, as these sample types involve targeting a specific sample associated with the target species (if not targeting

the species itself), it could be argued that this should be referred to as 'non-invasive' sampling (Lefort *et al*. 2015). Some eDNA researchers have argued just that, with as strict a definition as 'genetic material obtained directly from environmental samples (soil, sediment, water etc.) without any obvious signs of biological source material' (Thomsen and Willerslev 2015). This definition therefore does not classify community samples of e.g. bulk samples of insects (Zhou *et al*. 2013), gut samples for dietary analysis (Schnell *et al*. 2010), or non-invasive samples from visible sources such as faeces (Bohmann *et al*. 2011), etc as 'eDNA'. (see Figure 1.2). Bulk DNA is DNA obtained from community samples targeting particular organisms, such as from plankton collected with a plankton tow or large-size organisms scraped from rocks or collected in grabs (Darling *et al*. 2017).

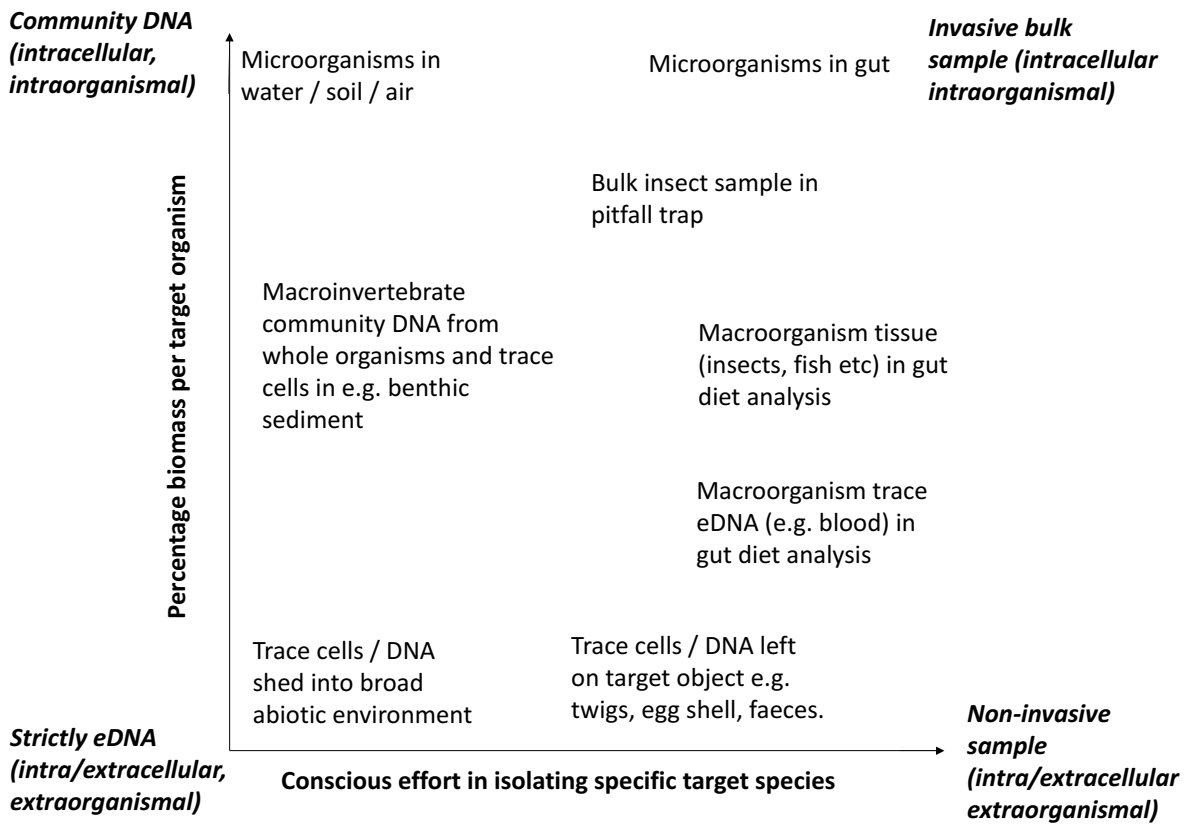


Figure 1.2. Plot of DNA sources. Examples of study topics which can be considered either 'Strictly eDNA', 'Community DNA', 'Invasive bulk sample' or 'Non-invasive sample' depending on the degree of conscious effort in isolating specific target species, (e.g. searching for egg shells or faeces) and percentage biomass per target organism (e.g. entire bacteria, or trace cells of a fish).

In this strict sense, eDNA is thought to be a combination of trace amounts of whole cells (intracellular DNA) and DNA fragments (extracellular DNA) (Turner *et al*. 2014b) shed into

the environment by organisms which are no longer present, and can then be detected by sampling the environment alone. Extracellular DNA from destroyed cells has usually degraded into small fragments (Beebee, 1991), whereas intracellular DNA comes from cells or organisms present within the sample, and is more likely to be high quality (Creer *et al*. 2016).

Whilst eDNA refers to the source of DNA, metagenomics, metagenetics and metabarcoding refer to the approach from which the analysis is performed, the main concept of which is analysis that transcends individuals (Greek 'meta' means 'transcendent', 'after', 'among', or 'beyond' as in metazoan: all multicellular animals). As barcoding is the study of barcodes, genetics is the study of genes, and genomics is the study of genomes, then metabarcoding, metagenetics and metagenomics can on a simple level be thought of as the study of all or many barcode genes, all or many genes, or all or many genomes, respectively. Metagenomics is defined as "the functional analysis of environmentally derived DNA" by Creer *et al*. (2010), who also defined metagenetics as "the large-scale analysis of taxon richness via the analysis of homologous genes". Handelsman (2009) defines metagenetics as the "application of mutant analyses in a community context" and suggests that whilst genetics and genomics deal with single organisms, metagenetics and metabarcoding provide a parallel with metagenomics, and both apply to analysis of a multigenome unit, or community. Taberlet *et al*. (2012b) draw particular attention to the definition of DNA metabarcoding as "high-throughput multispecies (or higher-level taxon) identification using the total and typically degraded DNA extracted from an environmental sample (i.e. soil, water, faeces, etc.)". This multispecies identification from metabarcoding meaning the mass amplification of a specific marker from many different DNA molecules, from different cells or individuals, rather than the mass amplification of entire genomes, or focusing on genomic function, as is the case with metagenomics. Although the field of metagenomics, metagenetics, and metabarcoding, has until recently been considered applicable only to microorganisms (i.e. intraorganismal eDNA), the concept of these meta-approaches is being applied to samples of eDNA for the analysis of multiple macrobial organisms (i.e. extraorganismal eDNA) through massively parallel technologies and microarrays. The advantage of macrobial over microbial metagenetics is that the number of taxa is considerably smaller, and species boundaries are more reliably understood (Lodge *et al*. 2012).

When to use these terms, or others such as 'ecometagenetics' (Porazinska *et al*. 2010), 'ecogenomics' (Chariton *et al*. 2010) or 'metasystematics' (Hajibabaei *et al*. 2011), is

therefore somewhat contentious, possibly due to the recent emergence of a variety of mechanisms and situations in which they could apply (Handelsman *et al*. 2009; Eisen, 2012; Watson, 2014; Esposito and Kirschberg 2014). Regardless of semantics, the exciting message (introduced in Figure 1.3) is that these approaches are now used with respect to macrobiota, opening a breadth of new information for our understanding of species, communities, and ecosystems. For the purpose of this thesis, the concept of environmental DNA will be discussed in its wider sense, to include the broad mutual approaches to extraction, amplification, sequencing, and analysis of samples, whilst focusing on aquatic eDNA.

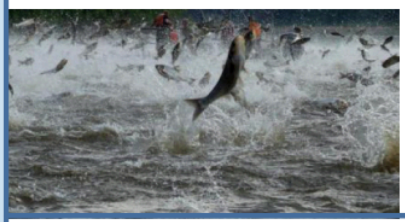
Through the use of eDNA (A) it is possible to obtain sequence information from the environment without isolating the target species first, which may detect species where traditional sampling has failed, **(B)** studies that necessitate rapid or multiple species detection are possible and ideally suited, **(C)** combined with 2nd Generation Sequencing, thousands or millions of sequences can be produced simultaneously to analyse species diversity.

(A) Sampling. Many species may be detected simultaneously.

Primers can be designed to amplify short fragments of degraded DNA (80-250bp) of one, or many target species using species-specific primers; or as many species as possible using universal primers. Often, mitochondrial markers such as Cyt B or COI are used as barcodes.

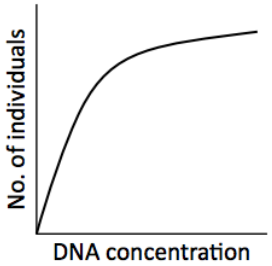
(B) Applications. Monitoring rare or invasive species, abundance estimates or studies on ecosystem processes are possible through the use of eDNA.

As eDNA methods are rapid and cost effective, studies aiming to detect invasive species such as

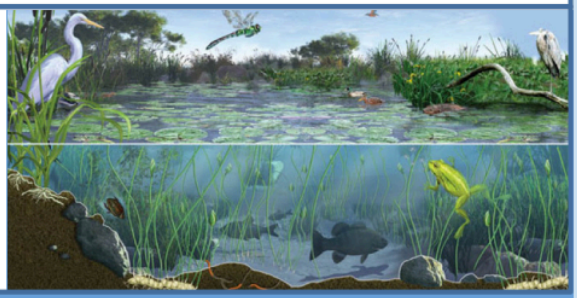


Asian Carp in the Great Lakes are particularly amenable to using eDNA.

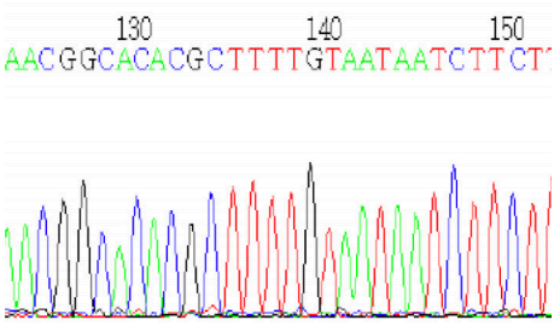
Studies have shown eDNA concentration to be directly related to number of individuals in mesocosms and natural ponds, but many issues still need to be addressed.



Data derived from the repeated sampling of single locations that describe dynamic relationships between taxa and the environment could help identify the role of niche-based stochastic processes in shaping species distributions and abundance. This type of information allows researchers to ask questions related to ecosystem processes.



(C) 2nd Generation Sequencing and eDNA. Combining 2nd Generation Sequncing with eDNA allows thousands of sequences to be analysed.



The use of 2nd Generation Sequencing allows in depth analysis through a variety of sequencing methodologies that are not possible with standard sequencing, such as the addition of tags to amplicons (when samples are pooled) to track which amplicons come from what sample; the generation of thousands of sequences at once which increases the reliability and scope of analysis; and the ability to sequence information in a much more cost-effective manner.

Figure 1.3. Sampling, applications and sequencing of eDNA. Summary of (A) the concept of environmental DNA (eDNA), (B) promising applications of eDNA, and (C) the advantages of combining eDNA with second-generation sequencing.

1.4 What can eDNA be used for?

As technologies have improved, the ability to detect tiny quantities of eDNA has increased dramatically (see Table 1.1), from identification of single species (Ficetola *et al.* 2008), to the detection of many species within a community (Thomsen *et al.* 2012b; Schnell *et al.* 2012; Anderson *et al.* 2012; Cannon *et al*. 2016), to exploring population variation (Sigsgaard *et al*. 2016; Stat *et al*. 2017). These studies cover a breadth of environments now more readily accessible to researchers when compared to traditional sampling. Studies that use environmental DNA in its strictest sense have mostly focused on proof of concept, however, there appears to be three overarching themes emerging for the use of eDNA: detection of species and biodiversity for conservation, biological research and monitoring of invasive species, and understanding ecosystem level interactions and patterns.

Until recently, it was thought that eDNA degrades so rapidly that only short fragment lengths are available for analysis in a similar way to aDNA, and subsequently eDNA amplicons have thus far been designed to be much shorter than those utilised in traditional molecular work. However recent studies have shown that in fact large fragments (Sigsgaard *et al*. 2016), entire barcoding genes (Deiner *et al*. 2016) and even entire mitogenomes (Deiner *et al*. 2017a) can be isolated from macrobial eDNA from a range of species, and that although eDNA is composed of short extracellular fragments, it can also be composed of whole intracellular DNA (Turner *et al*. 2014b).

Traditional detection of biodiversity may involve logistically challenging or expensive sampling methods such as casting nets, electrofishing, or even snorkel and SCUBA surveys (Jerde *et al*. 2011; Goldberg *et al*. 2011). However, recent work demonstrates the benefits of eDNA analysis. Access to challenging habitats such as the deep-sea (Corinaldesi *et al*. 2011; Guardiola. *et al*., 2015) or underground caves (Vörös *et al*. 2017) is possible with the use of non-invasive techniques, thereby minimising disruption to already fragile habitats and reducing disease transfer and stress to target species. Some examples of the different eDNA pipelines are given below in Figure 1.4 By using eDNA, researchers are offered a glimpse of the DNA from elusive and endangered species or undetected invasive species, particularly where they directly avoid conventional sampling methods.

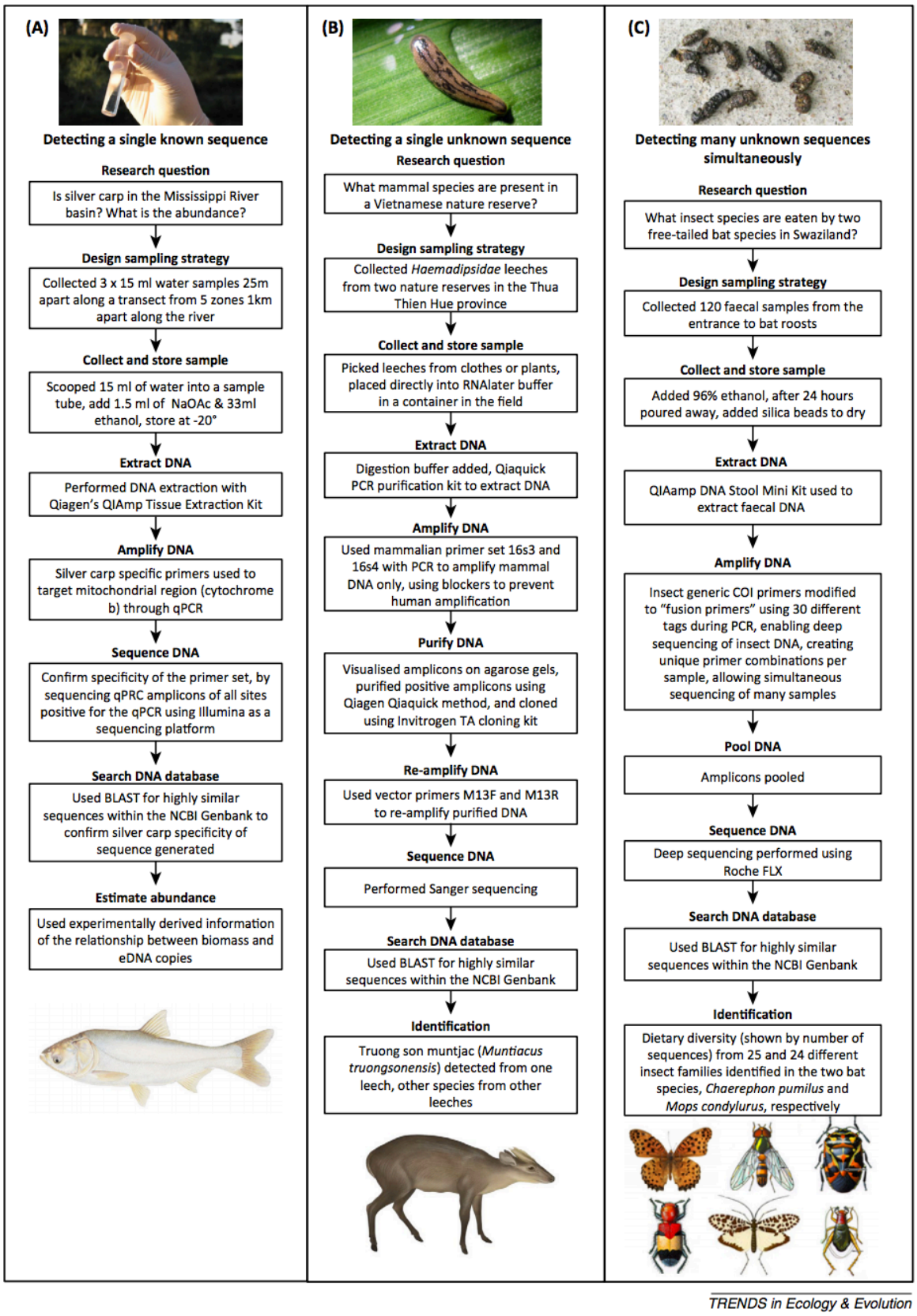


Figure 1.4. Exemplary environmental DNA (eDNA) case studies. These illustrate three research questions and the experimental procedures followed. (A) Detection of invasive Asian carp in a water sample (Jerde *et al.* 2011; Takahara *et al.* 2012). (B) Detection of mammal species in leech blood meals (Schnell *et al.* 2012). (C) Detection of insect prey in bat faeces (Bohmann *et al.* 2011).

Sample	Summary of study	Ref
French wetlands	Detection of invasive American bullfrog (*Rana catesbeiana*).	Ficetola *et al.* 2008.
Canals and waterways (USA)	Detection of invasive Silver Carp (*Hypophthalmichthys molitrix)* and Bighead Carp (*Hypophthalmichthys nobilis).*	Jerde *et al.* 2011.
Mountain streams in the USA	Detection of rare species: Rocky Mountain Tailed Frogs, (*Ascaphus montanus)*, and Idaho Giant Salamanders, (*Dicamptodon aterrimus).*	Goldberg *et al.* 2011.
Streams, ponds and lakes in Northern Europe	Detection of rare species: Common Spadefoot Toad (*Pelobates fuscus*), Great Crested Newt (*Triturus cristatus*), European Weather Loach (*Misgurnus fossilis*), Eurasian Otter (*Lutra lutra*), Large White-Faced Darter (*Leucorrhinia pectoralis*) and Tadpole Shrimp (*Lepidurus apus*). Also analysed eDNA concentration with relative abundance.	Thomsen *et al.* 2012a.
Streams in Spain	Detection of the chytrid fungus *Batrachochytrium dendrobatidis*, likely to be a primary cause of amphibian population declines.	Walker *et al.* 2007.
Forest pond water in Japan	Detection of multiple mammal species including Sika Deer (*Cervus nippon*), House Mouse (*Mus musculus*), Grey Red-Backed Vole (*Myodes rufocanus*), Raccoon (*Procyon lotor*), Brown Rat (*Rattus norvegicus*) and Long-Clawed Shrew (*Sorex unguiculatus*)	Ushio *et al.* 2017.
Seawater in the western Baltic	Detection of rare species: Harbour Porpoise (*Phocoena phocoena)* and Long-Finned Pilot Whale (*Globicephala melas*).	Foote *et al.* 2012.
Seawater near oil rigs in Qatar	Population variation analysis from eDNA detection of Whale Sharks (*Rhincodon typus*).	Sigsgaard *et al*. 2016.
Cave sediment in New Zealand	Extinct biota identified from cave sediment in New Zealand, revealing two species of ratite moa and 29 species of plants from pre-human era.	Willerslev *et al.* 2003.
Snow in Italy	Grey Wolf (*Canis lupus*) DNA isolated from blood spots in the Italian Alps and Arctic Fox (*Alopex lagopus*) DNA isolated from footprints.	Dalén *et al.* 2007.
Soil from a zoo in Denmark	Vertebrate DNA identified in soil samples collected in a zoolo matched to the elephant and tiger inhabitants, respectively.	Andersen *et al.* 2012.
Browsed twigs	Detection of Moose (*Alces alces*), Red Deer (*Cervus elaphus*), and Roe Deer (*Capreolus capreolus*), from saliva up to 24 weeks later.	Nichols *et al.* 2012.
Salt licks in Borneo	Detection of six endandered species: Bornean Orangutan (*Pongo pygmaeus*), Bornean Banteng (*Bos javanicus lowi*), Asian Elephant (*Elephas maximus*), Sunda Pangolin (*Manis javanica*), Sambar Deer (*Rusa unicolor*) and Bearded Pig (*Sus barbatus*).	Ishige *et al.* 2017.
Bromeliad water in Trinidad	Detection of the Golden Tree Frog (*Phytotriades auratus*) in their microhabitat of the Tank Bromeliad (*Glomeropitcairnia erectiflora*).	Torresdal *et al*. 2017.
Air	The presence of genetically modified organisms was detected from samples containing low levels of pollen.	Folloni *et al.* 2012.
Household dust	Detection of more than 600 unique arthropod genera inside 732 homes, including dust mites, cockroaches, and parasitic wasps.	Madden *et al.* 2016.

Table 1.1 Examples of the wide range of eDNA applications.

1.5 Advantages of aquatic eDNA as an assessment tool

One of the most well researched and widely implemented sources of eDNA in biodiversity assessment is that from water, which will from herein be the focus of this thesis. Aquatic sampling targeting eDNA has the potential to be implemented in routine biomonitoring (Baird and Hajibabaei 2012; Aylagas *et al.* 2014; Aylagas *et al.* 2016), assessment of conservation priorities (Minamoto *et al*. 2012; Yoccoz *et al*, 2012a; Barnes and Turner, 2016) and fisheries management (Evans and Lamberti, 2017; Hansen *et al*. 2018). As with other types of eDNA, whether or not the benefits of aquatic eDNA analysis are sufficient to enable uptake for management will depend crucially on the cost-effectiveness of any such new tools and the ease and efficacy of the approach. It is noteworthy that, as with the introduction of DNA barcoding *sensu stricto* (Hebert *et al*. 2003), which aimed to complement the Linnaean system of taxonomy, aquatic eDNA will most likely exert a pervasive impact through its integration with existing approaches rather than necessarily replacing them. For example, by evaluating the use of eDNA in detecting marine mammals, Foote *et al*. (2012) showed that conventional static acoustic monitoring devices that recognise echolocation were more effective in detecting the Harbour Porpoise (*Phocoena phocoena*), but eDNA better detected the rare Long-Finned Pilot Whale (*Globicephala melas*), indicating that eDNA is best used in conjunction with conventional approaches, also confirmed by others (Thomsen *et al*. 2016; Shaw *et al*. 2016; Hinlo *et al*. 2017a; Kelly *et al*. 2017). Although it has been suggested that aquatic eDNA will not replace traditional sampling and taxonomic expertise, there are several advantages of using aquatic eDNA to generate information regarding biodiversity quickly and efficiently.

a) *Higher sensitivity* – Detection probability for rare species when using traditional approaches to species monitoring is particularly low in aquatic environments, where individuals are hidden below the surface (Jerde *et al*. 2011), and so eDNA methods provide a way to access DNA from these unseen individuals. Higher sensitivity also comes from improved precision, as objectively identifying individuals from DNA barcodes is more accurate than visual taxonomic identification based on diagnostic morphological criteria that may leave room for subjectivity. Even when taxonomic skills are excellent, it may be near impossible to distinguish between juvenile individuals of animal groups such as fish, and consequently may also be difficult to make reliable management decisions such as those regarding the control of invasive species (Darling and Mahon 2011). Early studies show that reliable detection of

animals from aquatic eDNA at very low densities or small body size is possible where they may elude traditional sampling methods. For example, Thomsen *et al*. (2012a), detected eDNA from eight ponds where the Common Spadefoot Toad (*Pelobates fuscus*) had not been recorded using conventional survey methods, and Dejean *et al*. (2012) detected American Bullfrog (*Rana catesbeiana*) eDNA from five ponds where visual encounter and call detection had not recorded bullfrogs (confirmed by expert surveys). Similar results demonstrating the increased sensitivity of eDNA relative to traditional monitoring methods, particularly when combined with metabarcoding, have since been confirmed by others (Jerde *et al*. 2011; Darling and Mahon 2011; Olson *et al*. 2012; Thomsen *et al*. 2012b; Ji *et al*. 2013; Pilliod *et al*. 2013; Smart *et al*. 2015; Matsuhashi *et al*. 2016; Olds *et al*. 2016; Deiner *et al*. 2016; Valentini *et al*. 2016; Evans *et al*. 2017b; Eiler *et al*. 2018, Boussarie-Bakker *et al*. 2018). Initially validated by testing against artificially assembled communities of e.g. plants (Hiiesalu *et al*. 2012) or bulk insect samples (Yu *et al*. 2012), it has been demonstrated that metabarcoding generates reliable, qualitative estimates of alpha and beta diversity (Fonseca *et al*. 2010; Yoccoz *et al*. 2012b; Ji *et al*. 2013). However, artificially assembled communities may not provide a translatable illustration of genuine eDNA samples in real ecosystems (an important factor to understand when informing management decisions) (Lawson Handley, 2015). Some studies have, however, found that metabarcoding data and standard biodiversity sets are highly consistent (Ji *et al*. 2013).

b) *Usable for non-experts* – Protocols and sampling kits can be developed to enable citizen-science approaches, such as that developed by the Freshwater Habitats Trust and partners (Spygen, ARC and University of Kent) in the UK. In 2013, this group tested for the Great Crested Newt (*Triturus cristatus*) with promising results (Biggs *et al*. 2014; Biggs *et al*. 2015). Subsequently, this group completed the first ever national eDNA survey as part of the PondNet project in 2015 with 316 ponds, and again in 2016 with more than 550 ponds, and further sampling in 2017 (Freshwater Habitats Trust, 2017). As a result, Natural England has now approved eDNA analysis for monitoring the great crested newt (GOV.UK, 2017), which is being implemented by advisory services such as ADAS for various planning consultancies (ADAS, 2017). Another example is the larger scale citizen science project currently employed by the University of California (CALeDNA), which aims to characterise aquatic sediment

samples in and around California to build up detailed and complex distribution maps, with samples stored over time to compare both spatial and temporal patterns (CALeDNA, 2018).

c) *More cost effective* – The ease of sampling, and higher level of throughput of samples that may be processed allows information to be generated more cost-effectively (Shokralla *et al*. 2012; Calvignac-Spencer *et al*. 2013), although for qPCR-based studies, this depends on the cost of primer/probe development and the number of samples (Smart *et al*. 2016; Qu and Stewart, 2017). Michelin *et al*. (2011) showed that eDNA survey costs were 2.5 times cheaper and 2.5 times less time-consuming when detecting the invasive Bullfrog. Evans *et al*. (2017b) found that eDNA analysis of Brook Trout was 67% cheaper than electrofishing. Lugg *et al*. (2017) demonstrated that eDNA was more cost effective than trapping when targeting platypus, especially when combined with site occupancy detection models. Further to these studies, if eDNA approaches incorporate metabarcoding and NGS of a high number of samples, costs will be most efficiently reduced.

d) *Rapid sample collection and generation of results* – due to the short sample collection and analysis time, information may be generated more rapidly than by conventional survey methods (visual, acoustic, etc.), allowing a swifter management response (Darling and Mahon 2011). Sampling time also links in to sampling cost, as for example in the case of the eastern hellbender salamander (Olson *et al*. 2012), the greatest saving was in person-hours; whereas, typically, large teams are required for traditional sampling by rock lifting, a single researcher can collect and filter water, also demonstrated by Dejean *et al*. (2011).

e) *Non-invasive sampling* – There is no risk of harming target species through the use of true eDNA (as opposed to e.g. metabarcoding of bulk samples of insect pitfall traps), compared to trapping, netting, electrofishing or using biopsy darts for aquatic macrobiota (Jerde *et al*. 2011; Goldberg *et al*. 2011). This improves animal welfare, and researchers need not necessarily go through the process of tissue sampling and the associated permit applications, particularly for CITES-listed taxa.

1.6 Diversity of methodology for analysis of eDNA

Workflows utilising eDNA may range from simple 'yes/no' answers using quantitative or real-time PCR (qPCR) or conventional PCR (cPCR) pertaining to an individual species with no gene sequencing involved, to metagenomic sequencing of thousands of species in parallel. With a diverse array of sampling, isolating/capturing, DNA extraction, primer optimisation, PCR protocols and sequencing available, it is of high priority to compare their efficacy and application under a range of biological and abiotic conditions (Lodge *et al*. 2012) as some studies have explored (Renshaw *et al*. 2015; Deiner *et al*. 2015; Shaw *et al*. 2016; Eichmiller *et al*. 2016b; Spens *et al*. 2017; Schiebelhut *et al*. 2016; Piggott, 2016; Minamoto *et al*. 2016; Williams *et al*. 2016; Roy *et al*. 2017; Hinlo *et al*. 2017a and 2017b; Clarke *et al*. 2017; and Katano *et al*. 2017). Most of these however, have focused on proof of concept or method development, and there are as of yet few standard protocols in place to answer a particular ecological question. Generally, eDNA concentration is low in aquatic environmental samples and therefore a capture method is required to concentrate eDNA for molecular analysis. A consensus sampling methodology would benefit long term monitoring as confounding variables may create bias in interpreting ecological information. For example, varying pore sizes of different filter membranes may give biased results as varying eDNA concentrations may only reflect different particle sizes (Turner *et al*. 2014b; Wilcox *et al*. 2015; Shogren *et al*. 2016), rather than abundance or biomass of individuals (Barnes and Turner, 2016). Larger pore sizes (up to 5 μm) can make it easier to filter turbid waters, and produce higher eDNA yield (Thomas *et al*. 2018), with the use of a pre-filter step (an initial filtration using a broad pore size filter before a second filtration with a more fine pore size filter) being particularly helpful to decrease processing time without compromising detection probability (Robson *et al*. 2016, Bálint *et al*. 2017; Li *et al*. 2018).

1.6.1 Isolation –The water type (e.g. clear mountain stream/turbid tropical lake) and size of the target sample (e.g., bromeliad water/lake water) or organism (e.g., plant/nematode/fish) dictates the approach and quantity of the sample to be processed before DNA extraction (Creer *et al.* 2016). Collecting small volumes (usually 15 mL) of water for ethanol precipitation (e.g. Ficetola *et al.* 2008; Dejean *et al.* 2012), or filtering larger volumes (usually 1-2 L) of water (e.g. Goldberg *et al.*, 2011; Wilcox *et al.* 2013) have been the main methods of isolation of aquatic eDNA, with filtering becoming the predominant choice (Rees *et al.* 2014; Goldberg *et al.* 2016). However, success has still been achieved using

centrifugation and ethanol precipitation approaches, such as Klymus *et al*. (2017b) who found the same number, or a greater number of species using ethanol precipitation compared to filtering approaches, possibly due to extremely turbid sample water. One study even used filtering of up to 100 L on site using a specialised filtration capsule (Envirochek HV 1 lm; Pall Corporation, Ann Arbor, MI, USA) and a peristaltic pump (Valentini *et al*. 2016), a similar approach was also then implemented by Civade *et al*. (2016). Samples are generally either collected by hand from near the surface (e.g. Jerde *et al.* 2011), or at depth using limnological water samplers (e.g. Eichmiller *et al.* 2016a) using a sterilized sample bottle or pumped via peristaltic pump (e.g. Goldberg *et al.* 2011). Different filter material has been used such as cellulose nitrate, glass fibre, and polycarbonate, as well as different water volumes (Fahner *et al*. 2018), and different storage techniques (Minamoto *et al.* 2016; Spens *et al*. 2017), such as RNAlater (Ishige *et al.* 2017), Longmire's buffer (Renshaw *et al*. 2015; Wegleitner *et al.* 2015; Williams *et al.* 2016), ethanol (Goldberg *et al.* 2011; Hundermark and Takahashi, 2018), cetyltrimethyl ammonium bromide (CTAB buffer) (Renshaw *et al*. 2015), benzalkonium chloride (Yamanaka *et al.* 2017), dry storage in silica gel (Bakker *et al*. 2017; Majaneva *et al.* 2018), freezing (Jerde *et al.* 2011; Takahara *et al.* 2015; Hundermark and Takahashi, 2018; Majaneva *et al.* 2018), or even Qiagen lysis buffer ATL (Majaneva *et al.* 2018)). Examples of studies using different filter materials, pore sizes, and storage mediums are more thoroughly listed in the supplementary material of Chapter 2.

1.6.2 Extraction – Extraction methods vary between different types of commercial kits or in-house protocols, with differing success across studies, between labs and within studies. For example, Amberg *et al*. (2015) compared the PowerWater® DNA Isolation Kit from MO BIO Laboratories Inc, and the DNeasy Blood and Tissue Kit from Qiagen, and found that the Qiagen kit outperformed the PowerWater kit, although there were varying results depending on which laboratory the extractions were performed at. Phase separation and precipitation methods (e.g. CTAB-chloroform and phenol-chloroform) seem to generally yield more DNA than silica column methods (e.g. MoBio and Qiagen kits (Renshaw *et al*. 2015; Deiner *et al*. 2015; Schiebelhut *et al*. 2016)), and give significantly different community structures from metabarcoding analysis (Djurhuus *et al*. 2017). Bead-beating of filters is sometimes used, and a recent comparative study suggests this step increases eDNA yield (Hundermark and Takahashi, 2018) but a consensus on the best practise for eDNA extraction for particular ecological questions has not yet been reached.

1.6.3 PCR – eDNA protocols have used both cPCR, and qPCR. Goldberg *et al*. (2011) tested different cPCR protocols, and found that the addition of the Qiagen Multiplex PCR kit improved detection in water filter samples over using Amplitaq Gold DNA polymerase and bovine serum albumin (BSA), although most eDNA studies have not incorporated this kit. Compared to cPCR, results from qPCR provide an rough comparative index of sample population size, as well as more sensitive detection (Lodge *et al*. 2012; Wilcox *et al*. 2013; Qu and Stewart, 2017; Williams *et al*. 2017), lower false positive rate (Amberg *et al*. 2015; Wilcox *et al*. 2015), and are more likely to amplify eDNA even in the presence of inhibitors that block amplification in cPCR (Amberg *et al*. 2015). Droplet digital PCR can also be used for quantification, and may be more cost efficient for many samples, improve sensitivity of detection, and reduce amplification bias compared to qPCR (Morisset *et al*. 2013; Nathan *et al*. 2014; Jerde *et al*. 2016; Hunter *et al*. 2018; Baker *et al*. 2018). However, when many species or entire communities are being targeted, multiplexing many samples using cPCR is necessary when pipelines include NGS metabarcoding. However, if the aim of an aquatic eDNA study is to detect several key species of importance, metabarcoding approaches may be wasteful if non-target sequence data is of no use. In this case, multiplexing qPCRs using species-specific primers has been suggested for simultaneous detection of multiple species from aquatic eDNA (Tsuji *et al*. 2018). PCR choice will therefore depend on whether the ecological question has to do with quantification, targeting a specific species, or analysing whole communities.

1.7 How does the probability of detecting species by eDNA vary?

Researchers and organisations employing eDNA approaches, along with the stakeholders, methodological developers, resource managers and policy makers, must be made aware of the current levels of uncertainty associated with eDNA. This is critical when eDNA methodology is being used to inform management or development decisions, such as those faced by local planning authorities responsible for enforcing environmental regulations with regard to planning developments and endangered species.

Water sampling illustrates the complexity of interpreting eDNA-based studies. Detection probability is likely to be dependent on the interplay between DNA release and DNA degradation (Dejean *et al*. 2011; Thomsen *et al*. 2012a) as well as a range of variables which behave differently across habitat types (Barnes *et al*. 2014; Goldberg *et al*. 2016;

Barnes and Turner 2016). These include: organism size (Klymus *et al*. 2015; Lacoursière-Roussel, 2016b), and/or biological activity (Bylemans *et al.* 2016; Dunn *et al*. 2017), season (Goldberg *et al*. 2011; Vervoort *et al*. 2012; de Souza *et al*. 2016; Buxton *et al*. 2017b; Sigsgaard, *et al*. 2017; Stoeckle *et al*. 2017; Uchii *et al*. 2017; Salter, 2018, Buxton *et al*. 2018; Collins *et al*. 2018), organism species density (Pilliod *et al*. 2013; Pilliod *et al*. 2014), DNA degradation and dispersal rates (Deiner *et al*. 2014; Wilcox *et al*. 2015; Jane *et al*. 2015; Goldberg *et al*. 2018) and DNA or cell sloughing/shedding rate (Lacoursière-Roussel, 2016b; Sassoubre *et al*. 2016); while host molecule density (e.g. discrete tissues varying in mitochondrial density) is likely also important. For example, it is speculated that animals such as crayfish which have hard exoskeletons, or turtles which have hard shells are harder to detect using eDNA methods (Raemy and Ursenbacher, 2018) as they are thought to excrete less eDNA than animals with softer, more slime-coated skin types such as amphibians and fish, which have been most studied using eDNA methods (Thomsen *et al*. 2012a; Tréguier *et al*. 2014; Barnes and Turner, 2016).

The life stage of a particular organism can also affect DNA concentration, as shown by Dunn *et al*. (2017) who found that the presence of crayfish eggs on ovigerous females increases eDNA detection. Aquatic eDNA degradation rate is likely to increase depending on numerous factors, including initial DNA fragment length (Jo *et al*. 2017), substrate type beneath the water body (such as topsoil, clay or sand) (Shogren *et al*. 2016; Jerde *et al*. 2016; Buxton *et al*. 2017a), increasing time after the target organism is removed (e.g. Goldberg *et al*. 2011, other examples discussed below), increased environmental temperature (Pilliod *et al*. 2014; Strickler *et al*. 2015; Eichmiller *et al*. 2016a; Lacoursière-Roussel; 2016b, Lance *et al*. 2017; Tsuji *et al*. 2017a), increased or decreased pH (Strickler *et al*. 2015; Lance *et al*. 2017; Tsuji *et al*. 2017b), increased exposure to ultraviolet light (Pilliod *et al*. 2014; Strickler *et al*. 2015), bacterial and/or fungal action (Matsui *et al*. 2001; Dejean *et al*. 2011; Lance *et al*. 2017), and DNAses. Salter (2018) demonstrated significant seasonal variability in the turnover of marine dissolved eDNA, which they found to be correlated with higher temperatures, subsequent enhancement of microbial metabolism, and low concentrations of bioavailable phosphate, resulting in increased microbial utilization of dissolved eDNA as an organic phosphorus substrate. However, Collins *et al*. (2018) found no statistical relationship between marine eDNA degradation and temperature variation between seasons. Other factors affecting the detection of DNA, which can sometimes be stochastic include suspended sediment particle size (Turner *et al*. 2014b; Wilcox *et al*. 2015; Shogren *et al*. 2016), water

body depth (Smart *et al*. 2015; Stewart *et al*. 2017; Minamoto *et al*. 2017), different water body surface points (Hänfling *et al*. 2016; Evans *et al*. 2017a), sediment load/turbidity (Williams *et al*. 2017), and water flow rates (Deiner *et al*. 2014). Compared to freshwater systems, marine systems present a more challenging habitat to sample due to the potential dilution of eDNA into expansive waters, salinity, tides and currents which are likely to make eDNA detection patterns much more complex (Thomsen *et al*. 2012b; Thomsen *et al*. 2016; Baker *et al*. 2018; Collins *et al*. 2018). Generally, it is thought that eDNA degrades faster in marine, rather than freshwater environments (Thomsen *et al*. 2012b; Sassoubre *et al*. 2016) although a recent study uncovered the opposite pattern (Collins *et al*. 2018).

Waterborne eDNA appears to yield near-real-time, local (in lentic waters), and reliable-but-noisy estimates of species presence. The fastest rate of decay in freshwater systems assessed to date is 1.2 h (Seymour *et al*. 2018), and in marine systems is 6.9 h (Sassoubre *et al*. 2016), with most estimates ranging from 10 to 50 h (Weltz *et al*. 2017; Collins *et al*. 2018). Estimates of aquatic eDNA persistence time once organisms are removed from their environment are highly variable between studies. The detection of eDNA has ranged from roughly one (Pilliod *et al*. 2014; Thomsen *et al*. 2012a; Thomsen *et al*. 2012b), two (Dejean *et al*. 2011; Thomsen *et al*. 2012a; Barnes *et al*. 2014; Pilliod *et al*. 2014) three (Goldberg *et al*. 2013), four (Dejean *et al*. 2011; Merkes *et al*. 2014), to seven (Strickler *et al*. 2015) weeks with amphibians, fish or molluscs in mesocosms, artificial ponds and laboratory aquaria with varying environmental conditions. Based on the above factors affecting eDNA degradation, eDNA will persist in dry, dark and cold environments better than wet, light and warm environments, hence why ancient environmental DNA studies have been so successful from these types of environmental conditions (e.g., Jørgensen *et al*. 2012a; Jørgensen *et al*. 2012b; Giguet-Covex *et al*. 2014; Willerslev *et al*. 2014) and why sampling from warm, bright, aquatic habitats (such as tropical lakes) therefore, is likely to only yield genetic information from very recent biological activity. Studies focusing on soil or lake sediments have found that detectable traces of plant and animal eDNA persist from a few years (Andersen *et al*. 2012) to millennia (e.g., Haile *et al*. 2007; Yoccoz *et al*. 2012b; Hebsgaard *et al*. 2009; Giguet-Covex *et al*. 2014) or even tens to hundreds of millennia (Suyama *et al*. 1996; Willerslev *et al*. 2007). Ancient or historic eDNA could, however, contribute a possible source of error for modern aquatic eDNA sampling if sediment is re-suspended in the freshwater or marine water column (Barnes and Turner *et al*. 2016). Water samples rather than sediment samples therefore, are more likely to accurately reflect the

timely presence of target DNA (Shaw *et al*. 2016), although it has recently been shown that both aquatic and sediment eDNA exhibit congruent seasonal fluctuations when targeting Great Crested Newt eDNA in ponds (Buxton *et al*. 2018).

Understanding the origin, state, transport, persistence and fate of eDNA in varying environments as discussed above is essential if this technique is to be rigorously applied to ecological questions, summarised in Figure 1.5. below. This aim will be better met by comprehensive, replicated sampling surveys across a range of species and habitats, drawing upon cross-disciplinary knowledge from e.g. microbiology and water quality monitoring. So far, the behaviour of eDNA particles appears to be inconsistent (Shogren *et al*. 2016) and complex (Jerde *et al*. 2016). For example, it has been demonstrated that lotic eDNA could travel a few km in a small stream to more than 100 km in a large river (Deiner and Altermatt, 2014; Pont *et al*. 2018), but is unaffected by stream bottom substrate (Jerde *et al*. 2016).

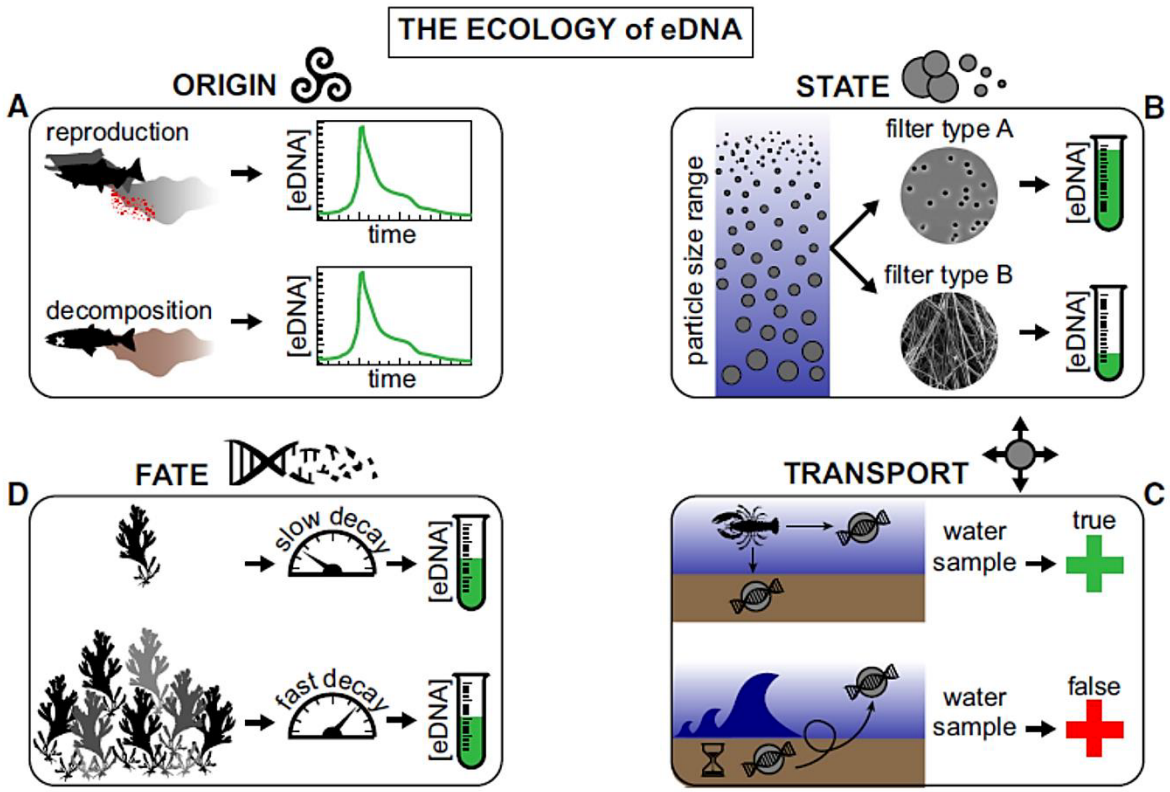


Figure 1.5. The ecology of eDNA (Barnes *et al.* 2016). The origin (A), state (B), transport (C) and fate (D) of eDNA are all defining factors in how eDNA information may be detected and interpreted.

1.8 Sources of uncertainty in eDNA information and solutions to minimise them

As with most technological advances, limitations remain, as do many challenges that need to be overcome. The potential implementation of eDNA approaches across disciplines indicates that it will be critical not only to sample, extract, and sequence eDNA in an efficient and cost-effective manner, but also to efficiently and reliably handle and analyse the typically massive data sets generated by next-generation sequencing platforms. eDNA studies would not only benefit from standardised methods for particular types of biodiversity-related questions, but also from highly standardised, international monitoring networks and cohesive multidisciplinary approaches that build on the traditional ecological and taxonomic knowledge, whilst integrating new genomic and e-technologies (Cristescu, 2014). Although eDNA methods applicable to a broad range of environments and their resident taxa are currently being tried and tested, work remains to ensure their reliability and repeatability (the variation in measurements taken by a single instrument or person) and reproducibility (whether an entire study or experiment can be reproduced in its entirety) (Kelly *et al.* 2014a). Similar to the related study of aDNA (e.g., Gilbert *et al.* 2005), eDNA approaches require rigorous standards and controls such as those outlined by Goldberg *et al.* (2016) and Ficetola *et al.* (2016), without which the information obtained might not only be noisy, but outright misleading. Errors or bias from molecular work could undermine overall confidence in eDNA for end users, and may have disastrous implications for management for conservation or invasive species if resources are unduly wasted (Ficetola *et al*. 2016). Considerable apprehension exists regarding the possible sources of uncertainty associated with eDNA, of which there are critical challenges for consideration where error can be introduced, providing a basis for future research to address, which may inform best practice solutions. Eliminating false positives (type I error: eDNA detected where target species is not present) remains a major challenge for eDNA studies, as the mere presence of eDNA does not necessarily indicate the presence of the relevant organism. False negatives (type II error: eDNA not detected where target species is present) are also problematic. Discussed below are challenges and solutions in relation to avoiding false positives and negatives at each step of an environmental DNA based experiment.

1.8.1 a) Challenge: experimental design for field sampling – Various factors discussed above are likely to determine the effectiveness of eDNA surveillance. In addition to the relationship between eDNA release and degradation, external sources of eDNA present a problem.

Dispersal of eDNA (in particular for air or waterborne eDNA) or contamination may result from the addition of DNA from other sources within the target environment (Dejean *et al*. 2011; Dejean *et al*. 2012), such as tributaries into a major river, ballast or bilge water discharge (Egan *et al*, 2015; Ardura *et al*, 2015b), sewage and wastewater (Martellini *et al*. 2005), excrement from animals that prey upon the target species, or dead target organisms (Darling and Mahon, 2011; Jerde *et al*. 2011; Merkes *et al*. 2014). A single sample from one site may not accurately represent local biodiversity due to the low probability of capturing all eDNA sequences at one time, and so it may take multiple samples to capture a particular DNA sequence (Andruszkiewicz, *et al*. 2017).

1.8.1 b) Solution: It is pertinent to understand how to correctly sample environments to capture the representative biodiversity within a given system, encompassing factors such as water depth or surface position, volume for filtration, and number of sample, extraction and PCR replicates. Experimental design, and interpretation of eDNA results should be carefully considered, and robust quality control implemented. Pilot sample collection should first be undertaken to test the logistics and efficacy of the sampling protocol (Goldberg *et al*. 2016). Negative field controls should be collected alongside experimental samples to ensure contamination does not occur in the sampling or transport phase (Goldberg *et al*. 2016), as for example, performed by Jerde *et al*. (2011) who ran 1L of deionised water through the filter apparatus between filtration of different samples. Field equipment, supplies and personnel should be kept separate from areas of high copy number DNA (i.e. PCR laboratories) prior to sampling (Goldberg *et al*. 2016). Equipment including boots, boats or field apparatus should be sterilised thoroughly. Most eDNA studies have used 10% bleach for around 10 minutes for sterilisation, such as Jerde *et al*. (2011), who reported no contamination in all blank samples. Recent guidelines, however, suggest that sterilisation should ideally be implemented with a 50% commercial bleach solution, or preferably, equipment not reused at all (Goldberg *et al*. 2016). Openly reporting contamination issues, particularly in metabarcoding studies, is important for the progress of eDNA science, as a recent study has done (Pont *et al*. 2018).

Quantity of sample, number of spatial replicates, and number of temporal replicates determine the strength of the evidence, with increasing confidence from a single positive sample, to multiple positive samples from a single trip, to repeated trips with positive samples, and repeated trips with positive samples over different time points (Jerde *et al*. 2011; Goldberg *et al*. 2016). If samples are immediately filtered and stored on site, or filtered in e.g. a car during transportation between sites, eDNA concentration can be best preserved

(Yamanaka *et al*. 2016a). Generally, the larger the amount of sample e.g. greater volume of water, the better. Mächler *et al*. (2016) recommend filtering at least 1 L of water for aquatic eDNA studies aiming to detect macroinvertebrates, although this is likely a recommendation specific to this system, with other aquatic eDNA samples perhaps needing more, or less, depending on the eDNA concentration and target organism. The type of sample used when assessing aquatic ecosystems will produce varying results. Shaw *et al*. (2016) showed that when sampling both from the water column, and the sediment surface of the benthic layer, eight species were detected from water samples and only three from sediment samples. This is also a matter of logistics, and so sample volume and number of samples should be considered within the trade-off between confidence in results and available survey time and budget (Smart *et al*. 2016).

Control samples can also be taken from adjacent areas where target species are known to be absent (Jerde *et al*. 2011; Ficetola *et al*. 2015) to allow further confidence in results, through non-amplification of these adjacent samples. Risk assessment of target eDNA emanating from other sources should be undertaken, including the presence of dead organisms (Goldberg *et al*. 2016). Repeated temporal sampling will provide a partial solution to the inability of eDNA to differentiate live and dead organisms, and control for eDNA left behind after a target organism is no longer present, as only live species that are permanently present will still be detected in repeated temporal samples. Another solution in differentiating between dead and live organisms is through the use of environmental RNA (eRNA), which is only detectable from live organisms (Pochon *et al*. 2017). How a sample is collected e.g. what filter type to use, and where to complete this step (i.e. field or lab) as well as whether to do multiple sample steps (i.e. a 'pre-filtration' step as in Turner *et al*. (2014b)) should all be considered depending on the target organism and environment in question (Goldberg *et al*. 2016).

1.8.2 a) Challenge: experimental design for molecular analysis – Detection tools must be highly sensitive and specific to avoid both false negatives and false positives respectively. False positives are a particularly problematic issue in environmental samples (especially if ancient) which contain low amounts of short fragment size DNA and typically require many PCR cycles to amplify (Ficetola *et al*. 2015). Contamination can also occur through trace DNA from laboratory surfaces which carry over into new reactions, or even extraction, PCR and sequencing chemistries (Darling and Mahon, 2011; Dejean *et al*. 2011; Goldberg *et al*.

2016). In simple presence/absence eDNA studies using cPCR or qPCR, false positives or negatives may also occur due to PCR primers and probes that do not have a high enough level of specificity, and allow the detection of "lookalike" non-targets (Dejean *et al*. 2011; Dejean *et al*. 2012; Darling and Mahon 2011; Wilcox *et al*. 2013). False negatives may occur from insufficient sensitivity or failure of methods to perform as expected (Darling and Mahon, 2011). For example, in metabarcoding approaches, PCR or primer bias may mask DNA of low quantity and over-amplify higher quantity DNA, which may skew the relative abundance of communities, leading to false negatives of certain rare DNA sequences (Bik *et al*. 2012; Cristescu, 2014; Elbrecht & Leese 2015; Piñol *et al*. 2015). False positives or negatives may also occur due to PCR errors such as 'tag jumping' in which unique tag sequences added to universal primers jump between samples, making it impossible to distinguish between samples (Schnell *et al*. 2015). It has also recently been suggested that DNA extracts from aquatic eDNA samples should not be pooled before sequencing, as these limits the detectability of rare sequences, particularly when targeting fish (Sato *et al*. 2017).

1.8.2 b) Solution: Molecular assays must be carefully designed and validated from pilot sampling prior to experimental activities getting underway, taking into account what extraction kit, PCR set-up, library preparation kit and sequencing approach to use (Goldberg *et al*. 2016). Both repeatability and reproducibility should be demonstrated for all assays (Darling and Mahon *et al*. 2011; Dejean *et al*. 2011). Potential inhibition of samples should be tested by either adding a foreign DNA and a matching assay to all samples (internal positive controls or mock samples) (Goldberg *et al*. 2016; Thomsen *et al*. 2016) or by creating a qPCR dilution series (e.g. Agersnap *et al*. 2017) from which an observed quantification cycle shift of >3 cycles is considered evidence of inhibition (Hartman, Coyne & Norwood, 2005; Goldberg *et al*. 2016). Inhibition can be removed either by diluting samples, or using a PCR inhibitor removal kit (e.g. Williams *et al*. 2017), both of which, however, may result in the loss of target DNA. To ensure specificity, *in silico* testing of species-specific DNA-based probes and primers (such as comparing sequences to BLAST (Altschul *et al*. 1990), or using ecoPCR software (Ficetola *et al*. 2010), or PrimerTree (Cannon *et al*. 2016)) as well as *in vitro* testing of probes and primers against target *and* non-target tissue-derived DNA should be standard procedure (Dejean *et al*. 2011; Darling and Mahon, 2011; Goldberg *et al*. 2016; Agersnap *et al*. 2017), and genetic distances should be reported (Jerde *et al*. 2011). This is particularly important when the outcome of a positive

result may be controversial, such as where management outcomes are likely to be affected such as in the control of invasive species. In these cases, positive PCR detections should also be sequenced to examine accuracy (Ficetola *et al*. 2008; Jerde *et al*. 2011; Thomsen *et al*. 2012a). However, it is not necessarily essential to design species-specific primers that do not amplify closely related species if these congeners do not have the same geographic distribution as the target species. This approach was employed by Goldberg *et al*. (2011) who designed species-specific primers for within the Rocky Mountains region only, or Dejean *et al*. (2011) who designed primers which amplified sturgeon congeners that were not found in their experimental ponds.

To reduce the incidence of false positives, assay design must include extraction and PCR blanks to the molecular workflow (Darling *et al*. 2011; Dejean *et al*. 2011; De Barba *et al*. 2014; Ficetola *et al*. 2015; Ficetola *et al*. 2016; Goldberg *et al*. 2016) allowing the explicit reporting of rates of false positives, and the formation of data filtering thresholds to remove background contamination in metabarcoding, as done by e.g. Thomsen *et al*. (2016) and Andruszkiewic *et al*. (2017). For conventional PCR, positive results observed in any negative controls render experimental samples suspect, and so should subsequently be discarded, unless quantification is the purpose of the study in which case very low amplification may be acceptable (Goldberg *et al*. 2016). However, if samples are for metabarcoding, and all samples are sequenced including controls, low level contamination is almost guaranteed, and can be bioinformatically filtered. Furthermore, the addition of endogenous positive controls using universal primers may distinguish between false negatives arising from method failure or reduced detection sensitivity (Ardura *et al*. 2015; Furlan and Gleeson 2016). If laboratory conditions are as sterile as possible, contamination indicated by the addition of extraction and PCR blanks should be minimised. To do so, most studies have used rooms specific to pre- and post-PCR activities; rooms dedicated to low-quantity DNA sources; rooms in which no DNA of the target species has been previously handled (Goldberg *et al*. 2011); or clean rooms such as used in aDNA studies (Dejean *et al*. 2012).

As well as including controls, increasing the number of technical replicates at the extraction and PCR step will enhance the reliability of data, as false negatives are less likely and false positives can be filtered out with proportionately lower thresholds (Cristescu, 2014; Ficetola *et al*. 2015; Leray and Knowlton, 2017), although the workload and costs obviously increase respectively. As eDNA can occur at such low concentrations, it is also important to use an appropriate volume of extract. Mächler *et al*. (2016) recommend screening at least 14

µL of extracted eDNA to reduce uncertainty in detections when targeting aquatic macroinvertebrates, although much like sample volume, this is likely to vary according to eDNA concentration and target organism. Ficetola *et al*. (2015) suggest at least six PCR replicates for eDNA metabarcoding when detection probability is around 0.5, or even eight PCR replicates if detection probability is lower than 0.5. This could easily be the case with studies aiming to screen unknown biodiversity present in samples.

One approach to increase the percentage of informative markers is to prevent non-target molecules from being enriched and sequenced by sequestering them with blocking oligonucleotides. This approach has so far mostly been used to exclude a relatively small set of contaminating molecules from being sequenced (e.g. as used in Vestheim and Jarman 2008; Schnell *et al.* 2010; Wilcox *et al.* 2014). However, as the amount of eDNA sequence data increases, it is conceivable that 'blocking libraries' for common environmental contaminants will be created. For example, blocking GC-rich molecules can reduce the amount of bacterial DNA sequenced in a library. It should however be noted that blocking primers have been shown to modify the proportion of non-target reads in metabarcoding (Piñol *et al.* 2015).

If laboratory set up is carefully considered, such as in the planning of metabarcoding in which primers should be tagged with identical forward and reverse tags and used only once per sequencing pool (Schnell *et al*. 2015), greater confidence in the molecular assay may be achieved.

1.8.3 a) Challenge: processing the data – Current barriers to the use of eDNA include the requirement for extensive training in molecular biology and subsequent genetic data analysis. There is a need for improved bioinformatics pipelines, statistical tools, and data sharing approaches if eDNA users are to accommodate the often-underestimated 'tidal wave' of data (Reichhardt, 1999) that it is now possible to produce from metabarcoding or metagenomic studies. The need for appropriate bioinformatics tools and centralised storage and infrastructure to accommodate robust algorithms has been noted for some time (Reichhardt, 1999; Bik *et al*. 2012; Cristescu, 2014). Although public databases such as NCBI, BOLD and Dryad do exist, the responsibility of storing original data largely falls to individual laboratories or genomic centres, whilst the cost of storing data remains more or less constant (Cristescu, 2014).

1.8.3 b) Solution: Global, coordinated efforts to integrate traditional approaches and effectively implement evolving technologies is underway, such as by the iBOL Project, the Atlas of Living Australia (Atlas of Living Australia, 2017), the Genomic Observatories Metadatabase (GeOMe) (Deck *et al*. 2017) and the Global Biodiversity Information Facility (GBIF) (GBIF, 2018). Biodiversity e-infrastructure will benefit from advances in 'Big Data' biodiversity informatics and e-research infrastructure such as these, allowing the integration of different taxon-level data within a phylogenetic and environmental framework, facilitating informed decision-making (La Salle *et al*. 2016).

1.9 OTU clustering for metabarcoding analysis

When combining eDNA with metabarcoding, many studies approach the assignation of sequences to species using the clustering of similar sequence variants into what are known as Operational Taxonomic Units (OTUs). OTU clustering techniques have typically been applied to microbial studies (Sogin *et al*. 2006) using 16S rRNA, but along with eDNA sampling, have since been applied to other groups of life from ancient and environmental samples. These OTUs are typically defined as a cluster of reads with 97% similarity, roughly approximating individual species. However, this may not be the case if a) a species has genes that are >97% similar, and so multiple OTUs are created for one species; b) a species may have paralogs that are <97% similar, and so multiple OTUs are created for one species; or c) artefacts such as read errors and chimeras can create spurious OTU clusters (Sokal, 1963; Sneath and Sokal, 1973).

There are three principal categories of species delineation: 1) clustering, 2) tree-based and 3) character based, with the first two being the dominant approaches used (Kekkonen *et al*. 2015). Clustering uses distance matrices, e.g. statistical parsimony networks such as jMOTU (Jones *et al*. 2011), Clustering 16S rRNA for OTU Prediction (CROP) (Hao *et al*. 2011), Automatic Barcode Gap Discovery (ABGD) (Puillandre *et al*. 2012), and Barcode Index Number (BIN) (Ratnasingham and Hebert 2013). These clustering approaches depend upon pairwise sequence distances between specimens to define the number of OTUs within a dataset (Kekkonen *et al*. 2015). Tree-based methods such as the Generalized Mixed Yule Coalescent (GMYC) (Pons *et al*. 2006), and Poisson Tree Processes (PTP) (Zhang *et al*. 2013), use a gene tree as input for the analysis, and may outperform clustering approaches in species assemblages lacking a 'barcode gap' (Zhang *et al*. 2013). The lack of a barcode gap is

usually linked to recently diverged species with little genetic diversification, and may in fact be an artefact of insufficient sampling across taxa (Wiemers and Fiedler, 2007), although examination of the width of the barcode gap with pairwise distances without *a priori* grouping can provide a preliminary approximation of divergence, and potential support in interpretation of results (Kekkonen *et al*. 2015). Character-based methods such as Character Attribute Organization System (CAOS) (Sarker *et al*. 2002), employ diagnostic base substitutions as a basis for assessments. The appropriate dissimilarity value to define OTUs is not only related to a specific method, but also to the sample complexity. Low complexity datasets need a higher dissimilarity threshold, whilst high complexity datasets need a stricter dissimilarity threshold, as the usual threshold of 3% often leads to under-estimation of OTUs (Chen *et al*. 2013). Other analysis tools exist as an alternative to cluster-based methods such as DADA2 (Callahan *et al*. 2016), which uses an error model to infer exact sample sequences which may vary by only a single nucleotide, or Swarm v2 (Mahé et al. 2015). This type of analysis could be beneficial for metabarcoding based upon short amplicons generated from universal primers amplifying e.g. the 12S Teleost primers from Valentini *et al*. (2016) which are less than 100 bp, and which therefore may not differentiate between fish species which differ by only one or few nucleotides (Stoeckle *et al*. 2017).

1.9.1 Normalization

OTU clustering requires a strategy to adjust for over or under represented OTUs, which may arise through PCR or sequencing bias, where relative abundance among samples could affect the resulting clusters (Molik *et al*, 2018). Normalization, or rarefaction, or transformation of read counts among samples within an OTU table is usually performed when analysing metabarcoding data (Molik *et al*, 2018). Variance stabilization such as R package DESEQ2 ((Love, Huber & Anders, 2014), used by Port *et al.* (2016)), cumulative sum scaling such as R package metagenomeSeq (Paulson *et al.* 2013), or subsampling-based normalization strategy (Aguirre de Carcer *et al.* 2011) are employed.

A correct method of controlling sequence quality to remove spurious sequences obtained through sequencing error, PCR error, lab contamination and so on is therefore important (Ficetola *et al*. 2016). Read trimming, filtering of artefacts/chimeras, reference database and/or de novo OTU generation, taxonomic assignment method and parameters as well as statistical analysis should all be carefully considered (Goldberg *et al*. 2016). Filtering data to remove sequencing artefacts may eliminate rare species, particularly when the

biomass of rare species is reduced (Zhan *et al*. 2014). As with most ecological data on species presence/absence and abundance, imperfect detection from eDNA data is unavoidable (Ficetola *et al*. 2015). Providing rules-of-thumb is impossible, but appropriate analyses can aid in better transformation of NGS reads into community information (Ficetola *et al*. 2016). Some studies opt to avoid OTU clustering all together, and individually blast all sequences (e.g. Thomsen *et al*. 2016). Species Occupancy Models (SOMs) can analyse species distribution when detection probability is lower than one, and estimate the number of replicates required for reliable interpretation of taxon absence (Pilliod *et al*. 2013; Schmidt *et al*. 2013; Schmelzle & Kinziger, 2016). Bioinformatic pipelines and programs have also been designed to improve estimation of diversity, taxonomic assignment, and statistical inference such as the Amplicon Pyrosequence Denoising Program (APDP) (Morgan *et al*. 2013), PRObabilistic TAXonomic placement program, 'PROTAX' (Somervuo *et al*. 2017), LULU (Frøslev *et al*. 2017), 'insect' in R (Wilkinson *et al*. 2018), the Mitochondrial Genome Database of Fish (MitoFish) and MiFish pipeline (Sato *et al*. 2018), or the use of informatic sequence classification trees (Wilkinson *et al*. 2018).

1.10 What barcoding markers are suitable for eDNA?

Many eDNA barcoding primers have been designed for the detection of specific organisms, or taxon groups, based on particular genes. Because eDNA samples may contain highly fragmented DNA, many universal barcoding primers (termed 'mini-barcodes') have been designed to target short fragments of 90-250 bp (examples shown in Table 2 below). Different gene regions vary in taxonomic coverage and species-resolving power, with specific taxonomic biases and imperfect estimates of taxon relative abundance (Creer *et al.* 2016). Ideally, metabarcoding markers should have sufficient taxonomic coverage to detect groups of interest, sufficient sequence divergence to resolve species, be conserved among individuals of the same species, indicate relative abundance of present taxa, be easy to amplify and create a short enough amplicon length to avoid sequencing error (Clarke *et al.* 2017). These can be used individually, or combined in a 'primer cocktail' of multiple primers at once (Ivanova *et al.* 2007). Mitochondrial or chloroplast genes present desirable molecular markers due to their uniparental inheritance, rapid mutation rates, multiple copies per cell, and ease with which conserved PCR primers may be designed for them (Handley, 2015). The COI gene is a popular choice for eukaryotes, as a previously 'agreed' region for standardisation of barcoding by molecular ecologists and conservation geneticists, who have

over time accumulated over 5.7 million specimens (BOLDSYSTEMS, 2017a) with barcodes freely available on the BOLD database (Ratnasingham and Herbert, 2007). Some studies have combined species-specific COI primers with environmental DNA analysis (Bronnenhuber and Wilson, 2013), whilst others have combined universal degenerate primers for a range of genes (Hänfling *et al.* 2016). This gene can provide an excellent marker choice, and it has recently been suggested that COI should be the standard barcode gene of choice for metabarcoding (Andújar *et al*. 2018). COI is extensively covered within DNA sequence reference databases, and it has a high degree of sequence variation. For example, COI resolved up to threefold more taxa to species level compared to 18S in a study of zooplankton assemblages by Clarke *et al*. (2017). However, COI is not always suitable in other cases. More conserved priming sites have been suggested for metabarcoding of particular taxa, as the protein-coding COI does not always contain suitably conserved regions for species discrimination (Deagle *et al.* 2014), such as within nematodes which are more often targeted using the 18S rRNA gene (Floyd *et al.* 2002; Powers, 2004).

For plants, which have low substitution rates of mitochondrial DNA, two plastid DNA regions '*rbc*L' and '*mat*K'; a gene '*trn* H – *psb* A'; and a nuclear ribosomal DNA region 'ITS' have been suggested as candidates for taxonomic assignment (Coissac *et al.* 2016; Fahner *et al*. 2016). Other popular gene regions include 12S and 16S for vertebrates, and ITS1 for fungi (see Table 1.2. for a small number of examples). Multiple primers targeting different regions are sometimes used in combination to increase species barcoding information, such as conducted by e.g. De Barba *et al.* (2014) when assessing diet composition of brown bears, by Shaw *et al.* (2016) when conducting a fish community assessment in rivers, or by Hänfling *et al.* (2016) when assessing fish communities in lakes. This multi-gene approach reduces taxonomic bias and increases taxonomic coverage (Alberdi *et al.* 2017). Combining many samples with many universal primers of different genes has been coined 'Tree of Life' (ToL) metabarcoding, by Stat *et al*. (2017) who combined nine primers targeting 18S, COI, 16S, trnL, and 23S genes, amplifying eDNA from 434 eukaryotic taxa from 38 phyla, 88 classes, 186 orders and 287 families.

Best practice involves evaluating barcodes according to certain criteria such as size, specificity, versatility, taxonomic resolution, understanding of the mode of evolution, and how comprehensive the taxonomic database is (Cristescu, 2014). Barcodes for commonly used metabarcoding markers are generally lacking from public databases, although COI is fairly well represented (Porter and Hajibabaei, 2018). It has recently been suggested that

classification of COI metabarcoding data could be improved by the use of the Ribosomal Database Project (RDP) classifier, which is faster than BLAST, and provides a measure of confidence for assignments at each rank in the taxonomic hierarchy (Porter & Hajibabaei, 2018). Primer design software such as Primer3 (Rozen and Skaletsky, 1999), ecoPrimers (Riaz *et al.* 2011) and PrimerMiner (Elbrecht and Leese, 2017) have been developed to aid in designing primers which take these factors into account.

Gene	~ Amplicon size (bp)	Taxon	Reference
COI	658	Vertebrates	Ward *et al* 2005
	313	Metazoa	Leray *et al*. 2013
	120-150	Eukaryotes	Meusnier *et al*. 2008
	100	Lepidoptera and fish	Hajibabaei *et al*. 2006
12S	163-185 and 55-75	Fish	Miya *et al*. 2015 and Valentini *et al*. 2016
	40-60	Batrachia	Valentini *et al*. 2016
	40-60	Enchytraeidae	Epp *et al*. 2012
	40-60	Aves	Epp *et al*. 2012
16S	100-120	Coleoptera	Epp *et al*. 2012
	54-60, 90, and 106	Vertebrates	Palumbi *et al*. 1996, Taylor, 1996 and Riaz *et al*. 2011
	178-228	Fish	Berry *et al*. 2017, Deagle *et al*. 2009
	90-213	Crustaceans	Berry *et al*. 2017
18S	336-423	Eukaryotes	Pochon *et al*. 2013
	240-420	Eukaryotes	Stat *et al*. 2017
ITS1	122	Aquatic plants	Gantz *et al*. 2018
	180-220	Fungi	Epp *et al*. 2012
	400-900	Vascular plants	Fahner *et al*. 2016
*rbc*L	400-900	Vascular plants	Fahner *et al*. 2016
	100-200	Land plants	Little *et al*. 2014
trnL P6-loop	40-60	Bryophytes	Epp *et al*. 2012
	40	Vascular plants	Taberlet *et al*. 2006
MatK	186	Aquatic plants	Gantz *et al*. 2018
23S	122-163	*Symbiodinium*	Santos *et al*. 2003; Manning & Gates 2008

Table 1.2. Examples of the variety of primers targeting different genes according to taxonomic group.

1.11 Challenges of barcoding and metabarcoding

There are various approaches to metabarcoding which use a combination of PCR and unique synthetic oligonucleotide sequences to label and multiplex samples during library preparation for next-generation sequencing, and subsequently reassign a sample to a sequence (Son and Taylor, 2011). These approaches namely consist of either a single PCR treatment or double PCR treatment (Figure 1.6). Contradictory terminology has been used interchangeably in the

literature to describe these unique oligonucleotides, which have been referred to as 'barcodes', 'tags' or 'indexes' resulting in inconsistency and confusion as to their exact function (O'Donnell *et al.* 2016). Within this thesis, these unique oligonucleotides will be referred to as 'indexes', and when combined with primers as 'index primers'. The unique oligonucleotides used to label individual libraries will be referred to as the 'library index', as per suggestion by O'Donnell *et al.* (2016).

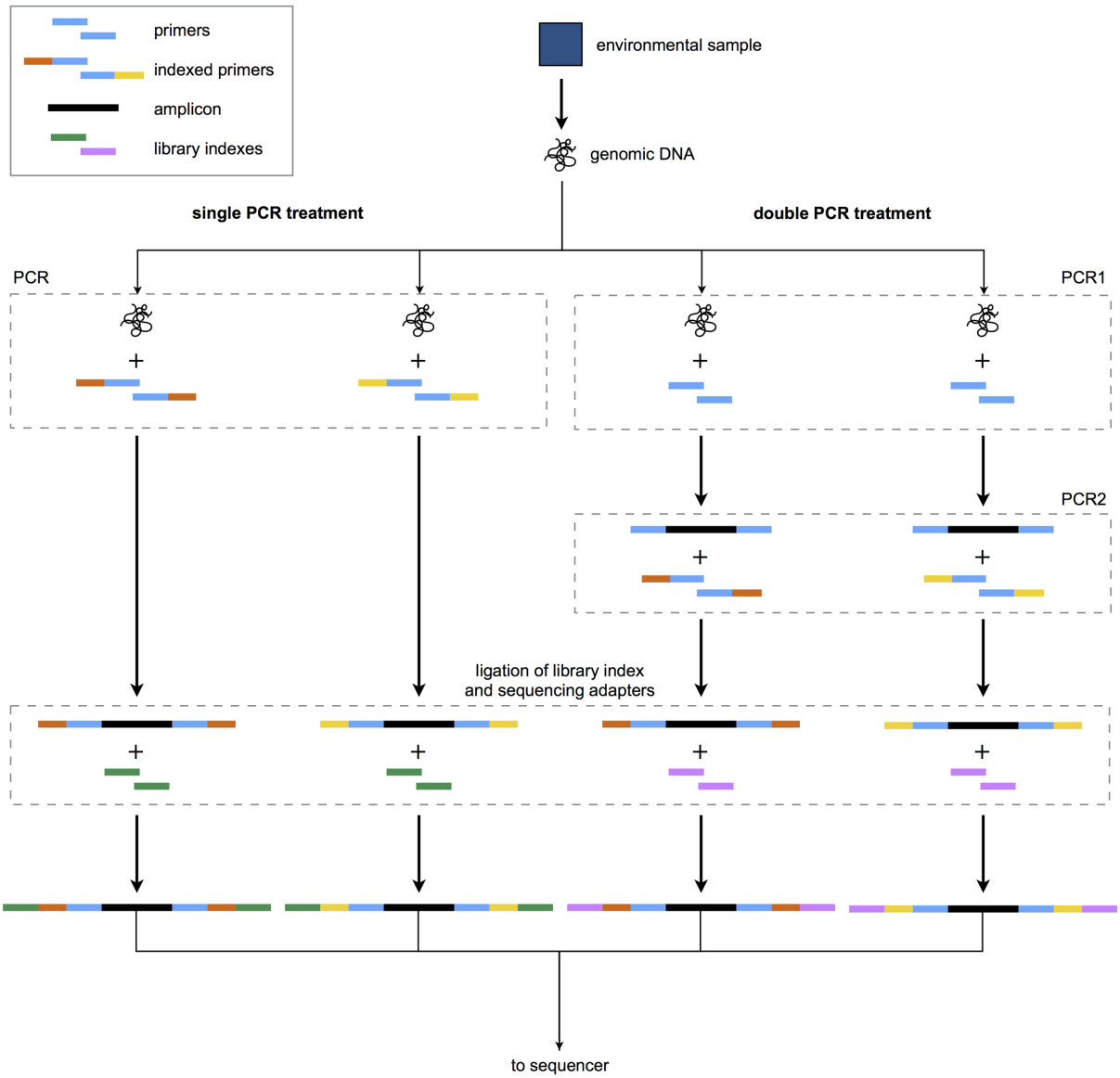


Figure 1.6. Single vs Double PCR treatment. Differently coloured primer and library indexes represent unique index sequences used to identify the sample origin of reads generated after sequencing on an Illumina MiSeq. Some studies use a single PCR treatment, whilst others use a double PCR treatment (O'Donnell *et al.* 2016)

For a single PCR approach, multiple unique index primers (Figure 1.5) are used to amplify a DNA extract, and then unique library indexes are ligated onto these sequences. For a double PCR treatment, a conventional PCR with unlabelled primers is performed, followed by a

second PCR to anneal the unique oligonucleotide labels. Mismatches between a primer sequence and template DNA reduces amplification efficiency of a PCR (Suzuki *et al.* 1996; Polz and Cavanaugh, 1998; Wintzingerode *et al.* 1997; Sipos *et al.* 2007), and when amplifying mixed templates, as with metabarcoding PCRs of eDNA, this can result in over-representation of template sequences which do not have mismatches (Suzuki *et al.* 1996; Wintzingerode *et al.* 1997; Piñol *et al.* 2014; Pinto *et al.* 2012). This can create inconsistencies in the relative abundance of OTUs when using a single PCR approach with index primers, compared to a double PCR approach using unlabelled primers for the first PCR (O'Donnell *et al.* 2016; Leray and Knowlton, 2017). This issue has been explored within the context of creating quantitative metabarcoding data by Piñol *et al*. (2018), who found that some primers pairs produced quantitative results reflective of the initial mock community used, whilst others did not. This study demonstrates how although quantitative estimates from metabarcoding can be roughly successful, the number of primer–template mismatches presents a challenge when attempting to use metabarcoding in a quantitative way when applied to the real-life variety of species richness and diversity observed in nature.

1.12 Abundance estimates using eDNA

A major opportunity provided by quantitative analysis of eDNA is to move beyond measures of the presence–absence of a species to its relative abundance in natural systems (Jerde *et al.* 2011; Minamoto *et al.* 2012). The ability to record not only how many species are present, but also how many individuals reside within any given habitat allows ecological queries to move from measures of species richness to species diversity. This yields advanced data for biodiversity and ecosystem monitoring, allowing the tracking of changes in ecosystems over time, observation of differences between habitats and ecosystems, and understanding the health of ecosystems. Indeed, the overarching question related to the next step for the use of eDNA is how it can be implemented for enumeration of individuals, and subsequently the creation of abundance estimates. A positive result alone from a natural aquatic environment can only indicate that at least one individual is or was recently present (Jerde *et al*. 2011). Although presence–absence measures can provide useful indicators of biological diversity, they are often insufficient to link rare species to persistence in a given habitat, or biological diversity to ecosystem functioning (Faust and Raes 2012). Rapid measures of abundance or biomass across time and space would be more informative and, importantly, could reveal seasonal shifts in factors such as microhabitat use for feeding and/or reproduction or refuge

use, as well as impacts of predation and competition. For example, Bylemans *et al.* (2016) analysed the relationship between bigheaded carp spawning and mitochondrial eDNA concentration. They demonstrated the use of nuclear, rather than mitochondrial markers, to detect fish spawning events in which spikes in nuclear eDNA concentrations were observed where no such spikes occurred from mitochondrial eDNA. Erickson *et al.* (2016) attempted to analyse the same question but found no such relationship.

There have been many attempts to relate eDNA concentration to either biomass or abundance, with inconsistent results. Some studies showed a strong correlation between the two whilst others showed weak or no correlation. Early studies positively correlated eDNA concentration from qPCR with broad categorical variables of high/low density of e.g. frogs in ponds (Ficetola *et al.* 2008), and Asian carp in different waterways (Jerde *et al.* 2011). This was later expanded upon by more refined abundance categories of e.g. numbers of American Bullfrog (*Rana catesbeiana*) tadpoles in experimental tanks (Dejean *et al*. 2011); number of Common Spadefoot Toad (*Pelobates fuscus*) and Great Crested Newt (*Triturus cristatus*) in experimental mesocosms (Thomsen *et al*. 2012a); number of Common Carp (*Cyprinus carpio*) (Takahara *et al*. 2012) or African Jewelfish (*Hemichromis bimaculatus*) (Moyer *et al*. 2014) in aquaria and experimental ponds; abundance of Wood Turtle (*Glyptemys insculpta*) in different rivers (Piliod *et al*. 2013); number of Lake Trout (*Salvelinus namaycush*) from catches in different lakes (Lacoursière-Roussel *et al.* 2016a); aquatic plant, Esthwaite Waterweed (*Hydrilla verticillata*) biomass (Matsuhashi *et al.* 2016) and abundance of a stream fish, Ayu, (*Plecoglossus altivelis)* (Doi *et al.* 2017b). The use of metabarcoding and next-generation sequencing of eDNA as a high-throughput means of obtaining measures of abundance across large scales and many taxa simultaneously has since been demonstrated (Kelly *et al*. 2014b; Elbrecht and Leese 2015; Klymus *et al.* 2017a), offering the promise of detecting cooperative and competitive relationships through robust tests of co-occurrence. Studies on this topic have found positive correlations between the number of sequencing reads and known community relative abundance of organisms from a range of environments, e.g. of bulk insect samples (Elbrecht and Leese 2015; Klymus *et al.* 2017a), fish from a marine aquarium (Kelly *et al*. 2014b), fish and amphibians in mesocosm experiments (Evans *et al*. 2016), freshwater fish in British lakes (Hänfling *et al.* 2016), Greenlandic deep-water marine trawl catches (Thomsen *et al*. 2016), and fish biodiversity from a large river (Pont *et al*. 2018).

However, these relationships have been calculated with a range of data points using a known density of individuals against eDNA concentration. The opportunity to estimate abundance based on concentrations of eDNA relies in part on the assumption that the release of eDNA from faeces, secretions, or tissues is correlated with the abundance or standing biomass of the respective individuals. This is likely to vary between life stages, individuals, species and habitat types (as discussed above) which could confound inferences about population size or biomass (Barnes and Turner, 2016). Confidence in eDNA generated relative abundance information would be improved by increased understanding of the persistence of eDNA in the wild from a broad range of climates and habitats, of how environmental factors affect eDNA concentrations, and of how accurately eDNA sequence copy numbers reflect the original composition of DNA in an environmental sample, and are not altered somewhere along the analytical pipeline. For example, PCR bias may lead to preferential amplification of some template sequences over others, and so the resulting diversity and relative abundance of the sequence reads may not necessarily reflect that of the community in the sample (Piñol *et al.* 2015; Bass *et al.* 2015).

1.13 Improving eDNA sequencing

Future eDNA studies are likely to take an increasingly metagenomic approach. Instead of PCR enriching a relatively small number of markers before sequencing, the eDNA extract will be sequenced in its entirety. If PCR is avoided completely, libraries have to be prepared directly from potentially highly degraded eDNA. Most existing library preparation protocols are optimised for high-quality DNA and are inefficient for highly degraded DNA (Knapp *et al.* 2012; Knapp *et al.* 2010; Gansauge and Meyer 2013). To overcome this limitation, eDNA methods can benefit from developments in the field of aDNA which routinely produces potentially relevant protocols in this regard (e.g., Knapp *et al.* 2010) such as recent progress in single stranded DNA library preparation from degraded DNA. Until the sequence output of second-generation sequencing platforms becomes sufficient to avoid informative marker targeting, enrichment methods are needed. Although PCR represents the basic option, hybridisation-based sequence capture offers an alternative (Liu *et al.* 2016; Wilcox *et al.* 2018). With an ability to target short molecules, under relatively permissive levels of mismatch (Taberlet *et al.* 2012a), such methods bypass major disadvantages of PCR enrichment.

Direct shotgun sequencing in e.g. metagenomic studies avoids potential taxonomic biases and can provide a complementary independent method to assess community alpha- and beta-diversity, and community functional genomic capability independent of the resolution of genetic markers, which often introduce bias (Cristescu, 2014; Creer *et al.* 2016). This approach avoids the biases and errors introduced by all target-enrichment strategies, such as tag-jumps observed in the PCR step of tagged primers for metabarcoding (Schnell *et al.* 2015). The power of Illumina-based direct shotgun sequencing of bulk insect samples was demonstrated by Zhou *et al.* (2013), with bioinformatic recovery of informative markers from the output. As sequencing costs drop and outputs increase, we might for the first time obtain directly quantifiable data representing the unbiased components of an eDNA extract. With the arrival of single-molecule sequencers (e.g., Pacific Biosciences (Ribeiro *et al.* 2012), Oxford Nanopore GridION™ and MinION™ (Schneider, and Dekker, 2012)) that remove the need for amplification during library build, these benefits will increase yet further. Progress in eDNA-based functional genomics will likely benefit from shotgun sequencing, especially if public metagenomic databases improve so that taxa, genomes and gene functions can be assigned (Creer *et al.* 2016).

1.14 The future of eDNA

eDNA is on the brink of making significant contributions to our understanding of invasive species, community and ecosystem processes underpinning biodiversity and functional diversity, and wildlife and conservation biology. Recent years have seen rapid improvements in sequencing technologies and we are only beginning to see the associated opportunities for eDNA research. It is enticing to imagine the possibilities that eDNA could unfold, if advances in molecular ecology, bioinformatics, and sequencing technologies continue to accelerate.

The main advantages of eDNA are rooted in its autonomous nature; with a reduced need for human taxonomists, ecologists, or biologists, sampling can access inhospitable environments (such as the Arctic, the deep sea, or even other planets), target elusive species, provide a vast reduction in labour costs and an increase in speed. Automated mechanical sampling of eDNA similar to that of oil spill-sampling buoys or military sonobuoys has already been put into action, with the ability to extract DNA, perform qPCR, and transmit data back via satellite (Preston *et al.* 2011), and robotically navigate habitats using unmanned aerial vehicles (UAVs; drones) (Ore *et al.* 2015; Doi *et al.* 2017c), or remote control boats

(Spyboat, 2017). Custom made integrated sampling systems have recently been created, such as the ANDe environmental DNA sampling system (Thomas *et al*. 2018) which uses a portable pump within a backpack, integrating sensor feedback, a pole extension with remote pump controller, custom-made filter housings in single-use packets for each sampling site and on-board sample storage.

If such eDNA automated sampling is combined with new technologies and a range of other complementary data in the future, the potential for our understanding of biodiversity and ecosystem processes may be greatly enhanced. NGS sequencing technology, or technology currently being developed by Oxford Nanopore Technologies to sample, upload via USB, and analyse DNA using the handheld MinION™ opens a world of possibilities for eDNA sequencing, the technology for which is decreasing in cost allowing an increase in sequencing throughput and data richness (Coissac *et al.* 2016). For example, the MinION™ was recently used to test samples for Ebola in Guinea (Quick *et al.* 2016), with results generated in just 15-60 minutes. When combined with human or robotic sampling (Ore *et al.* 2015; Doi *et al.* 2017c) targeting environments of interest, analysis of eDNA, and the remote upload of information via smartphone or satellite, it could be possible to create a network of live biodiversity assessment. Bohan *et al.* (2017) suggest that ecosystem changes could be monitored on a global scale, at high temporal and spatial resolution, using relative abundance of OTUs generated by NGS sequencing of eDNA, combined with machine learning methods. The authors suggest this type of information could be used to reconstruct ecological networks and interactions, with automated sampling uploading such information to 'the cloud'. This type of accurate abundance data would provide a potential framework for global ecosystem network prediction and enable the development of ecosystem-wide dynamic models (Faust and Raes 2012). If additional information was overlaid, such as water depth, hydrological or other environmental movements, temperature, pH, indicator biomolecules such as environmental RNA or proteins, or habitat information, such as the current ongoing project to map the Earth's surface in 3D (Amos, 2012), it could be possible to identify the origin and state of eDNA. For example, RNA degrades faster than DNA, and is indicative of active gene transcription, making it more likely to show the presence of metabolically active cells and is thus a better indicator of live, rather than dead, organisms (Poulsen *et al.* 1993). This subject was recently explored by Pochon *et al.* (2017) who compared eDNA and eRNA from the same samples, and recommend that only OTUs that are present in both eDNA and eRNA data should be interpreted as evidence of live organisms. As well as live biodiversity assessment

networks, ecosystem-wide dynamic models, and mapping the ecology of eDNA, it has been proposed that Earth observation data may be connected to biodiversity and ecosystems through interpolating biodiversity point samples and building continuous landscape maps of species distributions, which may then draw on known data associated with these species (Bush *et al.* 2017).

1.15 Freshwater ecosystems of Southeast Asia

'Freshwater' is defined as water with a very low dissolved solids content (around 1000 mg l^{-1} of dissolved solids (American Meteorological Society, 2012)), although some freshwater environments such as river estuaries may extend out into the ocean whilst some isolated inland water bodies may be highly saline. The development of human society has significantly relied on freshwater ecosystems, with the birth of the rich and civilised early empires occurring in river valleys, such as the Egyptians of the Nile, the Romans of the Tiber, and the Mesoamericans of the Amazon (Scott, 1989). River basins provided fertile soils to grow crops and graze livestock; plentiful waters to catch fish; riverine forests to harvest timber and hunt wild game; as well as drinking water, transport and the opportunity for spiritual and cultural traditions (Scott, 1989). Today, we would label these inherent elements 'ecosystem services' or 'ecosystem goods' to place monetary value on the processes and resources provided by the natural world for conservation purposes, such as water supply, regulation and purification, control of infectious organisms, fisheries, game hunting, tourism and recreation (Kottelat and Whitten; 1996, Costanza *et al*. 1997).

A common feature of freshwater ecosystems is the intimate bond between these resources and processes, and biodiversity. Although freshwater ecosystems only occupy 0.01% of the Earth's water, and 0.8% of the Earth's land-surface, they are estimated to contain around 126 thousand plant and animal species, equivalent to roughly 9% of all described species (Balian *et al*. 2008; Dudgeon *et al*. 2006). The total number of freshwater vertebrate species excluding brackish fish is around 18,235: constituting 35% of all described vertebrates (Balian *et al*. 2008). It is currently estimated that there are roughly 34,515 species of fish globally (Eschemeyer and Fong, 2017), a number which has risen substantially since 2008 when estimates stood at 29,000 (Lévêque *et al*. 2008). Around fifty percent of these fish species inhabit brackish or freshwaters (Balian *et al*. 2008), indicating that freshwater ecosystems are exceptionally species rich, although encompassing only a small component of the global aquatic realm, with ever growing species estimates as new studies emerge. Tropical freshwater ecosystems are particularly species rich, supporting over one million species worldwide which depend upon these habitats for their survival (Cumberlidge *et al*. 2009). These may be obligate freshwater inhabitants such as fish, semi-aquatic taxa such as frogs, or any species intrinsically linked to the hydrological processes and ecosystem interactions within their environment including birds, mammals and reptiles (Abell *et al*. 2008).

Figure 1.7. Southeast Asia. (Google Maps)

The geographical region of Southeast Asia (SEA) (Figure 5.) consists of Brunei Darussalam, Cambodia, Indonesia, Lao People's Democratic Republic, Malaysia, Myanmar (Burma), Philippines, Singapore, Thailand, Timor~Leste (East Timor) and Vietnam (United Nations Statistics Division, 2012). When Alfred Russel Wallace sailed between the volcanic shores and hiked into the humid forests of the Malay Archipelago in the 1850s, the influence of man had done little to corrode the ancient and flourishing biodiversity of Southeast Asia (SEA). Wallace's seminal book, *The Malay Archipelago* (1869), revealed the exceptional endemism of this region, and the stark division of species between the Asian and Australian continents on either side of what became known appropriately as Wallace's Line. One hundred and fifty years later, the Malay Archipelago encompasses most of modern day SEA, hosting four of the Earth's terrestrial biodiversity hotspots: Indo-Burma (Cambodia, Laos, Thailand, Vietnam and Myanmar), Sundaland (Brunei, Indonesia, Malaysia, Singapore), Philippines and Wallacea (Indonesia) (Myers *et al.* 2000) These biodiversity hotspots also contain many tropical freshwater ecosystems within 'freshwater ecoregions', categorised by Abell *et al*. (2008). Particularly noteworthy ecoregions include the Mekong river basin (running through Tibet, China, Burma (Myanmar), Laos, Thailand Cambodia and Vietnam); the Chao Phraya river basin (Thailand); the Sittaung and Irrawaddy river basins (Burma (Myanmar)) and large parts of Sumatra and Borneo (Abell *et al*. 2008). SEA ranks second globally (after the Amazon) for freshwater species richness, with the Mekong Basin and large parts of Malaysia and Indonesia considered noteworthy (Collen *et al.* 2014). It is the World's richest region for freshwater turtles (Buhlmann *et al.* 2009), and fish, crustacean, insect and molluscan diversity is particularly high (Balian *et al.* 2008; Kottelat 2013). The evolution of this extraordinary biodiversity must be appreciated within the context of the region's intricate tectonic and climatic evolution (Lohman *et al*. 2011; De Bruyn *et al.* 2014), characterised by over 300 million years of continental collisions (van

Oosterzee 1997; Metcalfe 2011) which influenced the creation of the wide-ranging topography, hydrology, geomorphology and consequently climate (Morley 2012).

The significant biogeographic barrier of the 'Wallace's Line' separates the Sunda Shelf to the west (where Indochina, Sumatra, Java, Borneo and Bali are found) and the Sahul Shelf to the east (where Sulawesi, Lombok, and Timor~Leste are found). The islands of the Sunda Shelf were previously a contiguous landmass known as Sundaland when sea levels were low enough during the middle Eocene (ca. 49-45 mya), and the Pleistocene (2.5 mya to 11,700 years ago). Southeast Asian ecosystems have experienced repeated and significant geographic reductions associated with the periodic submergence of the Sunda Shelf between cycles of Pleistocene glacial and interglacial periods (Woodruff 2010). This repeated range compaction is thought to be an instrumental force in causing the colonization and subsequent diversification contributing to the hyperdiverse SEA communities that we see today (De Bruyn *et al.* 2014). Although most of SEA's biodiversity in islands such as Java has arisen through the accumulation of immigrants and *in situ* diversification, within-area diversification and subsequent emigration are the principal characteristics typifying Indochina and Borneo's biota in particular, which have been described as 'major evolutionary hotspots for Southeast Asian biodiversity' (De Bruyn *et al.* 2014). As current climate and geography are typical of only ~3% of the last 2.7 million years, the biota of SEA is currently in a refugial state, in which they occupy only 50-75% of their maximal Pleistocene extent (Woodruff 2010).

Although freshwater ecosystems are incredibly species rich, mounting evidence suggests that freshwaters are the most threatened ecosystem in the world, with roughly double the rate of biodiversity loss than terrestrial and marine environments, recorded between 1973 - 2000 (Kottelat and Whitten 1996; Saunders *et al*. 2002; Jenkins *et al*. 2003; Collen *et al*. 2014; CBD 2013). The WWF (2012) suggests that the tropical freshwater Living Planet Index has declined by 71%, a pattern that is particularly poignant in Southeast Asia, as shown in Figure 1.8. A, B. Threats include direct habitat alteration, over harvesting of aquatic animals (especially fish), pollution, invasive species and anthropogenically induced climate change. Global extent of wetlands decreased by ~50% during the 20th Century (Hails *et al.* 2008). Wetland loss is certainly higher in SEA than globally (Rowley *et al.* 2010), where most remaining wetlands have been converted to rice paddy fields, reservoirs, canals or storm drains. Sea-level rise will impose additional threats through further reductions in land area and an associated increase in the refugial state for SEA taxa. The ~646 million humans in

SEA require food, water, energy, consumables and living space, which threaten FEs through a range of interrelated activities such as oil and gas extraction, hydropower creation, agricultural development and urban expansion. Projections forecast populations in SEA to rise to 797 million by 2050 (Worldometers 2018), putting these services under increasing demand, and escalating habitat loss through river impoundment, urbanization, deforestation and land-use change. Threats to freshwaters occur at the physiochemical, trophic and habitat level, and can be split into five main categories: water pollution, flow modification, habitat degradation, over-exploitation, and species invasions, upon all of which environmental change acts (Dudgeon *et al.* 2006).

Conservation of freshwater ecosystems (FEs) is often overlooked, despite freshwater biodiversity declining faster than terrestrial or marine biodiversity since 1970 (Dudgeon *et al.* 2006; Collen *et al.* 2014). Freshwater conservation strategies are of critical importance in densely-populated regions such as SEA, where high rates of habitat loss and species extinction (Myers *et al.* 2000; Collen *et al.* 2014) coincide with manifest risks to human water security (Vörösmarty *et al.* 2010). As the global human population, sea-level, and temperatures rise, it is inevitable that threats to freshwater ecosystems will intensify. An increase in frequency of extreme weather events, combined with economic growth and a tendency towards development of coastal cities will exacerbate the effects of anthropogenic change in SEA, as evinced by events reported in the media since 2015. For example, forest fires ravaged Indonesia, intensified by the drainage of Bornean wetlands; one of the most severe El Niño weather events recorded in 50 years caused widespread drought; saline intrusion crept up the Vietnamese Mekong; and construction began on the US$3.5bn Xayaburi Dam on the Mekong mainstream in Laos. Areas such as SEA have some of the highest levels of biodiversity in the world, but are thought to be understudied due to the low per capita gross domestic product (GDP), low level of English speakers, their past or present experience of civil or international conflict, and their geographical distance away from countries hosting biodiversity databases such as the Global Biodiversity Information Facility (GBIF) (GBIF, 2018), usually hosted in the U.S.A or Europe (Amano and Sutherland 2013). Biodiversity surveys are also more comprehensive in affluent countries, with a longer history of research, which could also bias species distribution estimates, as shown by the location of barcode entries in the International Barcode of Life Data Systems (Ratnasingham and Herbert, 2007) (Figure 6, C). The existence and extinction of some species, particularly that of small or cryptic organisms may therefore simply go unrecorded (Brook *et al.* 2003).

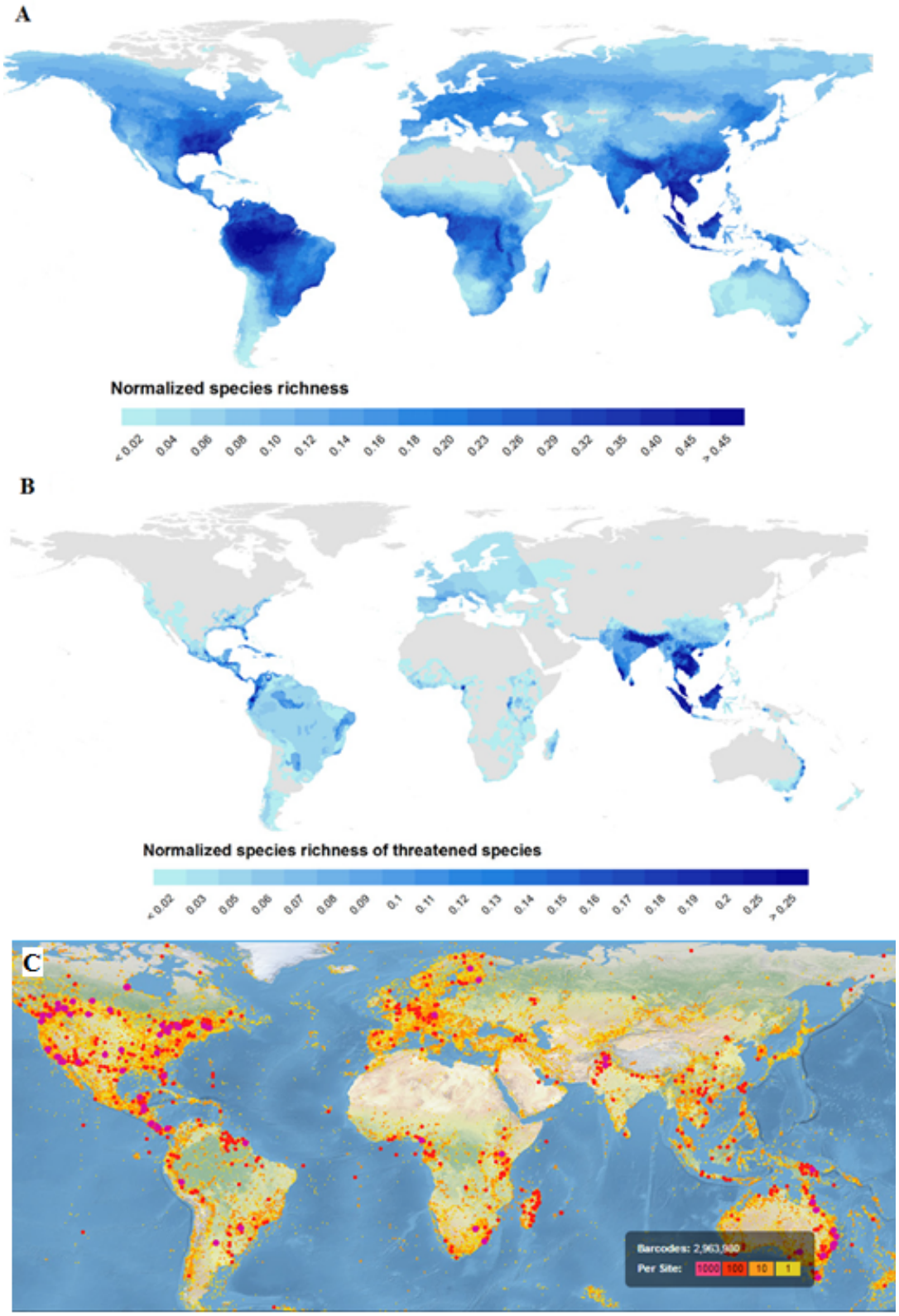


Figure 1.8. Global species richness compared to threat and BOLD database entries. A: normalized global freshwater species richness from 0 to 1. B: normalized global freshwater species richness of threatened species (extracted from Collen *et al*. 2014) C: global barcode entries to BOLD, pink = 1000 per site, red = 100 per site, orange = 10 per site, yellow = one per site. (BOLDSYSTEMS 2017b).

1.16 Significance of eDNA for Southeast Asia

The use of environmental DNA is particularly amenable to sampling of aquatic environments, which are among the most logistically difficult habitat types to sample using conventional methodology, hindered by the complexity of the topography and vegetation in streambeds and riparian areas, water turbidity and flow rates (Goldberg *et al*. 2011). Kottelat and Whitten (1996) divided Southeast Asia's freshwater ecosystems into habitat types of: 'springs, hill-streams, headwaters and rapids'; 'freshwater swamp forests and small streams in lowlands and foothills'; 'large rivers'; 'riverine lakes and flood plains'; 'estuaries'; 'lakes'; 'marshes and swamps'; 'peat swamps, black water streams and black water lakes'; 'caves and aquifers'; and 'artificial freshwater habitats'. It can be assumed that eDNA would behave differently in all of these freshwater habitat types, for example, moving quickly downstream from the source in a light, clear, cold 'spring, hill-stream, headwaters and rapids' habitat type, compared to moving slowly if at all in a darker, more turbid, warmer 'peat swamp, black water stream and black water lake' habitat type. In the wild, aquatic eDNA detection and degradation is likely to be complex, depending on climate variables, water body type and habitat variables as discussed in section 1.7. This is particularly relevant to Southeast Asian freshwater ecosystem types which are highly variable, and where DNA degradation is likely to occur faster, due to higher temperatures and microbial activity (e.g. Pilliod *et al*. 2014; Strickler *et al*. 2015; Eichmiller *et al*. 2016a; Tsuji *et al*. 2017a). In Indonesia and Malaysia, the combination of deforestation, the drainage of wetlands and conversion into agricultural land reduces the buffering capacity of rivers, creating higher peak flows, and lower base flows, resulting in higher risks of flooding and drought, as well as increased concentrations of suspended solids, resulting in higher levels of turbidity and reduced photosynthesis (Yule, 2004; Asian Development Bank, 2016). Indeed, the water quality of lakes and rivers in Indonesia is poor, with over 50% of water quality parameters not meeting the norms for water quality Class I (water that can be used as standard water for drinking purposes) (Asian Development Bank, 2016). The use of eDNA sampling could be amenable to understanding the impact of these threats on freshwater biodiversity. Three key areas of influence for aquatic eDNA application in SEA are 1) monitoring of invasive species, 2*)* understanding ecosystem level processes and patterns, and 3) monitoring for conservation management. Applying eDNA methods to address challenges within these topics could have major benefits for environmental protection, fisheries monitoring and management, or fishing and wildlife tourism in Southeast Asia.

1.16.1 Monitoring of invasive species

Invasive species present one of the most significant, inadequately controlled, and least reversible of threats to biodiversity and global homogenisation. They are 'highly noxious', dominating an area once they become established (Helfman, 2007), and spread becoming abundant (Kolar and Lodge, 2001), threatening biological diversity (Species Survival Commission, 2000). Other terms include non-native, nonindigenous, introduced, alien, exotic, transplanted, translocated, allochthonous, invasive, feral, and biological pollutant (Helfman, 2007). Such species may result in catastrophic effects for native freshwater ecosystems through competing with, predating on or transmitting disease to native species (Schneider *et al*. 2016), as well as causing eutrophication, reducing biodiversity, altering fire regimes, and destroying fisheries (Peh, 2010; Allen *et al.* 2012). In Southeast Asia, there is a substantial aquaculture industry as well as tourist game fishing, with many species being introduced either for their easily farmed meat (e.g. tilapia) (Guinée *et al*. 2010), or based on their size and attractiveness to fishermen. In Thailand, species such as the Giant Alligator Gar (*Atractosteus spatula*) from North America, the Silver Arowana (*Osteoglossum bicirrhosum*), and the Tambaqui (*Colossoma macropomum*), all from the Amazon, are introduced for sport fishing (Mega Fishing Thailand, 2017). Invasive species in SEA are introduced from a range of sources including aquaculture (e.g. Water Hyacinth (*Eichhornia crassipes*), Apple Snails (Ampullariidae sp.), and tilapia fish (e.g. *Oreochromis niloticus*), pest control (e.g. Mosquitofish (*Gambusia affinis*)) or the aquarium and pet trade (e.g. Armoured Catfish (Loricariidae sp.)) (Peh, 2010; Allen *et al.* 2012; Reid *et al.* 2013). In SEA, many examples exist where impacts of invasive species are observed at the physiochemical, trophic, and habitat level. For example, bioturbation and siltation have been caused by Common Carp (*Cyprinus carpio*), community composition has been altered by predation on fish by snakehead fish species (Channidae sp.), and habitat structure was impacted by Water Hyacinth (*Eichhornia crassipes*) and Floating Fern (*Salvinia natans*), which prevent movement of fishing boats and cause fishing net entanglement. Invasive species are more successful in degraded habitats (Allen *et al.* 2012), and so their impact will likely be compounded as deforestation, industrial agriculture and global temperatures increase (Peh, 2010).

The development of eDNA tools for application in monitoring invasive species has been one of the best studied aspects of eDNA, with most studies using qPCR methods (Table 3), although metabarcoding approaches have also recently been applied to search for non-

Species and native country	Habitat where eDNA found	Reference in date order
American Bullfrog (*Rana catesbeiana*) native to North America	Natural wetlands in France	Ficetola *et al*. 2008, Dejean *et al*. 2012
Asian Bigheaded Carp (*Hypophthalmichthys nobilis*) native to East Asia and Silver Carp (*Hypophthalmichthys molitrix*), and China / Eastern Siberia respectively	Lakes in the USA	Jerde *et al*. 2011, Mahon *et al*. 2013, Turner *et al*. 2014a
North American Bluegill (*Lepomis macrochirus*) native to North America	Ponds in Japan	Takahara *et al*. 2013
New Zealand Mudsnail (*Potamopyrgus antipodarum*) endemic to New Zealand	River in the USA	Goldberg *et al*. 2013
Quagga Mussel (*Dreissena bugensis*) native to Ukraine, and Zebra Mussel (*Dreissena polymorpha*) native to southern Russia and Ukraine	St. Joseph Lake and ballast water, the Rhine river catchment, and Lake Winnipeg, all in the USA	Egan *et al*. 2013, De De Ventura *et al*, 2017, Gingera *et al*., 2017
Louisiana Crayfish (*Procambarus clarkii*) native to northern Mexico, and Southern USA	Ponds in the Regional Nature Park of Brière, France	Tréguier *et al*. 2014
Burmese Python (*Python bivittatus*) native to South and Southeast Asia	USA	Piaggio *et al*. 2014, Hunter *et al*. 2015
North American Wedge Clam (*Rangia cuneata*) native to the Gulf of Mexico	Baltic Sea	Ardura *et al*. 2015a
Chinese Giant Salamander (*Andrias davidianus)* native to China	The Katsura River basin in Japan	Fukumoto *et al*. 2015
Tiger Mosquito (*Aedes albopictus*), Asian Bush Mosquito (*Aedes j. japonicus*) native to Southeast Asia	Natural water bodies in seven European countries	Schneider *et al*. 2016
Ruffe (*Gymnocephalus cernua*) native to Europe and North Asia	Laurentian Great Lakes	Tucker *et al*. 2016
Rusty Crayfish (*Orconectes rusticus)* and the Signal Crayfish (*Pacifastacus leniusculus*) native to North America	Lakes in Michigan, and the Laurentian Great Lakes of the USA	Dougherty *et al*. 2016, Larson *et al*. 2017
Northern Pike (*Esox Lucius*), native to the Northern Hemisphere	Alaskan lakes	Dunker *et al*. 2016
Red Swamp Crayfish (*Procambarus clarkii*) native to Mexico	Rice paddy water, Honghe Hani rice terrace	Cai *et al*. 2017
Topmouth Gudgeon, (*Pseudorasbora parva*), native to Asia	Angling ponds in southern England	Davison *et al*. 2017
Signal Crayfish (*Pacifastacus leniusculus*) native to North America, and Narrow-Clawed Crayfish (*Astacus leptodactylus*) native to the Caspian Sea	Natural freshwater ecosystems in Denmark	Agersnap *et al*. 2017
Brown Trout (*Salmo trutta*) native to Spain and invasive Rainbow Trout (*Oncorhynchus mykiss*)	Streams in the Biosphere Reserve and Natural Park of Redes, Northern Spain	Fernandez *et al*. 2018
Wild Pig (*Sus scrofa*) native to Eurasia and invasive in the USA.	Artificial wallows in Mississippi, USA.	Williams *et al*. 2018

Table 1.3. Examples of studies using eDNA and qPCR to detect invasive species.

indigenous shellfish in Spain (Borrell *et al*. 2017). Although invasive populations may rapidly reach large population sizes, the initial propagules in an early invasion are low-density and subsequently difficult to detect (Barnes and Turner, 2016). eDNA therefore presents a useful solution for providing rapid and accurate information on species' distributions as an early-warning system, to assess the geographic extent of current invaders, and to alert regulatory authorities before the establishment of alien species.

By sampling sources of invaders in transit such as ship ballast water (Li *et al*. 2011; Mahon *et al*. 2012; Egan *et al*. 2015; Ardura *et al*. 2015b), ornamental fish transport (Collins *et al*. 2013; Roy *et al*. 2017), recreational fishing bait trade (Mahon *et al*. 2014; Nathan *et al*. 2015), or at port locations (Grey *et al*. 2018), invaders may be detected and management action taken before the potential invasives arrive at their destination. Indeed, eDNA methodologies have already demonstrated particular promise in this regard. The US Fish and Wildlife Service, for example, have implemented an eDNA-based approach to monitor invasive Asian carp in the Midwest, USA (Figure 3.), providing a labour- and cost-effective alternative to traditional large-scale sampling methods such as electrofishing and/or manual netting (Jerde *et al*. 2011).

The ability to detect an invasive species early, and respond quickly, is of paramount importance for their management (Ficetola *et al*. 2008; Lodge *et al*. 2012; Dejean *et al*. 2012; Jerde *et al*. 2011). Populations at low densities must therefore be detected before they become established, allowing a much greater chance of eradication. For assessing biosecurity risk, the mantra is 'an ounce of prevention equals a pound of cure' (Lodge *et al*. 2012); knowledge of exact species distribution contributes to this by allowing preventative measures to most effectively be put in place.

Without the tools to detect rare invasives, and consequently a lack of knowledge on which to base a management plan upon, actions can stagnate or fail to begin. However, with the use of new information and practises, quantitative procedures for risk analysis, and cost-effective diagnostic technologies amongst other solutions, the effectiveness with which managers can respond to such situations may be improved (Lodge *et al*. 2006). Nevertheless, managers have been slow to adopt eDNA detection tools in decision-making frameworks that have a direct impact on management responses, possibly due to the remaining susceptibility to error (Darling and Mahon, 2011). However, there are instances in which eDNA has been implemented in such approaches; most prominent was the use of eDNA in the detection of invasive Asian Carp species in North America (Jerde *et al*. 2011). In 2008, the U.S. Army

Corps of Engineers (USACE) entered into an agreement with the Centre for Aquatic Conservation at the University of Notre Dame to carry out a risk assessment, which included testing for invasive Asian Carp species within the Chicago Sanitary and Shipping Canal and the Great Lakes (Darling and Mahon, 2011). Silver (*Hypophthalmichthys molitrix*) and Bighead (*Hypophthalmichthys nobilis*) Carp DNA was detected from environmental samples in areas previously thought to be absent of carp, north of where an electric barrier had been constructed to prevent carp dispersal, as far as Lake Michigan (Jerde *et al*. 2011). This discovery suggested that the need for management action to prevent invasions was much more urgent than previously thought, based on traditional sampling methods, and led to calls for a full separation of the Great Lakes and the Mississippi River basin, as well as for closure of the hydrological lock that leads directly to Lake Michigan. This led to the filing of a lawsuit in the US Northern District of Illinois to seek immediate action to prevent the invasion of Asian Carp, and the scientific community scrutinised eDNA studies and their reliability, with particular focus on the invasive carp in Mississippi (Darling and Mahon, 2011). These studies demonstrate the interest in eDNA for the management of invasive species, and the potential for monitoring invasive species such as those in Southeast Asia.

1.16.2 Understanding ecosystem level processes and patterns

The forces that threaten biodiversity may only be truly understood when the description of extant species and the mechanisms through which biodiversity interacts with the ecosystem are also understood. There is an urgent need for ecosystem level understanding to inform system-level response to accelerating anthropogenic impacts on Earth such as climate change, pollution and deforestation which are having huge impacts in Southeast Asia, and will have a knock-on effect for food security, emerging diseases, how to manage natural landscapes and how to tackle the spread of invasive species (La Salle *et al.* 2016). Realistic inferences and predictions about the impact of environmental change on extant biota depend increasingly on our ability to transcend boundaries among traditional biological hierarchies in the wild, extending from individuals to species, populations, and communities. Such an approach facilitates community eDNA analysis (Porco *et al.* 2010) simultaneously from across the kingdoms of life, including plants, animals, fungi, and bacteria. The ability of eDNA to move beyond targeted surveillance of a handful of species, to detecting multiple species simultaneously has great potential for community ecology and studies at the ecosystem level (Lodge *et al*. 2012). Building on microbial metagenomic approaches, eDNA sampling to

describe communities of organisms has evolved from studies of bacteria, to eukaryotic microorganisms, to macrobial life including invertebrates and vertebrates as well. Examples combining NGS and eDNA for the detection of multiple macrobial species (from a range of environments, not only aquatic) include the detection of communities of nematodes (Porazinska *et al*. 2010; Vervoort *et al*. 2012), earthworms (Bienert *et al*. 2012), plants (Yoccoz *et al*. 2012b; Fahner *et al*. 2016), amphibians (Bálint *et al*. 2017), fish (Thomsen *et al*. 2012b; Thomsen *et al*. 2016; Olds *et al*, 2016; Yamamoto *et al*. 2017), entire marine benthic metazoa (Leray and Knowlton, 2015), entire marine vertebrate communities including fish, marine mammals and birds (Port *et al*. 2016; Andruszkiewicz *et al*. 2017) and deep-sea marine octocorals (Everett and Park 2017). When combined with data derived from repeated sampling of single locations, the role of niche-based and stochastic processes in shaping species distributions and abundance, as well as life history activities could be identified (Haegeman and Loreau 2011). For example, a recent study demonstrated the use of aquatic eDNA metabarcoding in comparing sites affected by mining pollution, finding eDNA of previously unrecorded vertebrate species from mine polluted ponds (Klymus *et al*. 2017b). Harper *et al.* (2018) used presence-absence data of > 500 UK ponds to examine species associations between the Great Crested Newt (*Triturus cristatus*) and other vertebrates, and found that this species was significantly correlated with nine vertebrate species, and occurrence was broadly reduced where there were more fish species. Bakker *et al*. (2017) used aquatic eDNA from marine systems to detect 21 shark species, whose geographical patterns of diversity and sequence read abundance coincide with geographical differences in levels of anthropogenic pressure and conservation effort. Another recent study used fish bait to attract carp, and found that eDNA was up to 500x more concentrated at times of peak activity compared to a control environment of no bait (Ghosal *et al*. 2018). They also measured the hormone Prostaglandin F2α; PGF2α, which was correlated with higher eDNA concentrations, revealing the ability of baiting to increase not only the detection of aquatic eDNA, but also associated biological information with implications for assessing reproductive condition. This type of information could be beneficial for the conservation of Southeast Asian ecosystems by rapidly generating data concerning patterns associated with anthropogenic impact on biodiversity, as well as ecological fluctuations and animal behaviour.

The implementation of so-called ecosystem-based approaches (Clarke and Jupiter, 2010), which take a more holistic view than single-species studies, is particularly amenable to

eDNA, where trophic, energetic, and terrestrial–aquatic interactions can be detected and tracked. The field of parasitology for example, benefits from eDNA analysis which may aid in understanding host-parasite interactions, parasite communities, disease risk, the role of parasites in ecosystem processes as well as monitoring their spatial and temporal distribution between different life cycles for preventative measures (Bass *et al*, 2015). There have been several studies using aquatic eDNA to track a range of pathological organisms including parasites, bacteria and viruses. Gomes *et al.* (2017) predicted protozoan parasite outbreaks in fish farms, Hall *et al.* (2016) found a correlation between ranavirus found in pond eDNA and die-offs of the Wood Frog (*Lithobates sylvativus*), while Hartikainen *et al*. (2016) have used eDNA to assess myxozoan parasite diversity in aquatic environments which matched that from their vertebrate hosts. The amphibian chytrid fungus (*Batrachochytrium dendrobatidis*) is associated with massive population declines of amphibians in tropical countries, and has been detected in SEA since 2013, eDNA sampling has been effective in detecting this deadly fungus (Walker *et al*. 2007; Schmidt *et al*. 2013), and provides an effective technique for early detection and subsequent implementation of protection measures. In addition, trematode parasites infecting both amphibians and humans have been the topic of recent eDNA studies. *Ribeiroia ondatrae* known to cause morphological malformations including extra legs or the absence of legs in North American amphibians was recently targeted using eDNA with high specificity, consistently detecting as little as 0.001 pg through qPCR (Huver *et al*. 2017). *Opisthorchis viverrini* which can lead to cholangiocarcinoma in humans was also detected from ponds, rice fields, and rivers in Laos (Hashizume *et al*. 2017). These examples highlight the potential for eDNA in monitoring and managing the spread of parasites and disease for both animals and humans.

Complementary multidisciplinary approaches, such as combining aquatic eDNA with e.g. lake sedimentary aDNA and morphological analyses of micro- and macrofossils, show particular promise for elucidating the impact of changing climates on species and communities through time (Sarkissian *et al.* 2014; Jørgensen *et al.* 2012a; Anderson-Carpenter *et al.* 2011; Lejzerowicz *et al.* 2013; Sarkissian *et al.* 2014).

Key ecosystems underpinning plant biological production and carbon and nutrient cycling can be readily characterised using eDNA washed from root systems (Blaalid *et al.* 2012), generating insights into the dynamics of community structure and providing an ecological framework to investigate functional links among root-associated fungi, environmental variation and ecosystem diversity, and associated services. Such approaches

would be amenable to aquatic eDNA sampling of e.g. plants in riparian zones of rivers, littoral zones of lakes, mangrove forests or kelp forests.

Barberán *et al.* (2012) was among the first to link functional traits and biodiversity of microorganisms from DNA metabarcoding, yielding informative ecological markers by discriminating between marine ecosystems (coastal versus open ocean) and oceans (Atlantic versus Indian versus Pacific). Similar studies have used eDNA metabarcoding for ecotoxicology analysis using marine or freshwater benthic invertebrate communities, examining, for example, the effect of the antibiotic/antifungal agent, triclosan (Chariton *et al.* 2014), the effect of fish farming (Pawlowski *et al.* 2014), the effect of different land-use types (Saxena *et al.* 2015; Xie *et al.* 2016) or urbanisation (Kelly *et al.* 2016), with communities revealing a correlation between these drivers and their species richness. If studies such as these advanced to functional genomic analysis, it would be possible to identify adaptive or fitness-related loci, monitor loci related to stress events, or describe the molecular basis of inbreeding depression from environmental mixtures (Zepeda Mendoza *et al.* 2015).

Within the context of studies such as these, it has been suggested that eDNA metabarcoding will be transformative for biomonitoring or bioassessment (Baird and Hajibabaei, 2012), producing in the range of 10^3–10^4 species-equivalent operational taxonomic units (OTUs, encompassing all biota from microbes to metazoa) at a reasonable cost, and comparable biotic index (Aylagas *et al*. 2016). This has particular applicability for freshwater and marine ecosystems (Baird and Hajibabaei 2012; Aylagas *et al.* 2014; Aylagas *et al.* 2016) which are notoriously difficult to monitor using traditional methods. Environmental DNA for biomonitoring has proven comparably successful to traditional methods in aquatic environments (Mächler *et al*. 2014), and with increasing technologies and decreasing cost, is likely to provide a faster, more cost-effective and more efficient method for detection of a variety of indicator species including invertebrates, fish and algae. Indeed, there is talk of how to incorporate environmental DNA metabarcoding into standard monitoring for the European Water Framework Directive (WFD) to assess the "Biological Quality Elements" (BQEs), namely phytoplankton, benthic flora, benthic invertebrates and fish (Hering *et al*. 2018).

Aside from strictly aquatic eDNA monitoring, species monitoring for trophic and community interactions such as predator ecology, interspecific competition, or niche partitioning is particularly amenable to diet analyses or molecular scatology which share

common approaches to eDNA sampling in its strict sense (e.g., Clare *et al.* 2009; Razgour *et al.* 2011). Traditionally, diet analyses were performed either by directly observing what an animal ate or by collecting its faeces and examining prey fragments under a microscope. eDNA metabarcoding has provided an alternative or complementary approach, using faecal or other bodily extracts amplified with tagged universal primers (Binladen *et al.* 2007), making it more efficient and cost-effective to obtain diet information on a large scale (e.g., Bohmann *et al.* 2011; Deagle *et al.* 2009; Pegard *et al.* 2009; Soininen *et al.* 2009); reviewed in (Pompanon *et al.* 2012; Valentini *et al.* 2009). In addition to questions related to trophic interactions, dietary sampling provides insight for biodiversity monitoring. Because predators or blood-sucking insects feed on biodiversity, collecting either faecal material or the insect itself for molecular diet analysis can identify rare or cryptic species that traditional monitoring methods such as camera traps might miss. Recent studies include stomach-content analyses of parasitic invertebrates such as leeches (Schnell *et al.* 2012) (Figure 3), carrion flies (Calvignac-Spencer *et al.* 2013), mosquitoes (Kent, 2009), and ticks (Gariepy *et al.* 2012) to reveal their vertebrate hosts. In one case, Vietnamese terrestrial leeches of the genus Haemadipsa revealed the presence of the endemic Annamite Striped Rabbit (*Nesolagus timminsi*) that had not been detected despite monitoring the site for several thousand nights with camera traps (Schnell *et al.* 2012). In fact, leeches are currently being used to search for the highly endangered saola antelope in Vietnam and Laos (Saola Working Group, 2013), and provide a promising avenue for the monitoring of large vertebrates in Southeast Asia, with recent research exploring whether different leech species are more successful iDNA samplers than others (Drinkwater *et al*. 2018). These types of dietary approaches would complement aquatic eDNA sampling when assessing aquatic biodiversity and ecosystem processes, either through dietary analysis of e.g. fish guts (Leray *et al.* 2013) or capture of aquatic parasites such as the leeches within the Hirudidae family.

1.16.3 Monitoring for conservation management

Prompt assessment of precise species distributions is a vital requirement for conservation management (Magurran 2013; Dejean *et al*. 2011), and so the development of methods that improve detection probabilities is of high conservation priority. By their nature, species of most conservation concern are most often difficult to study due to their rarity and regulations on their sampling, handling, and transport of tissue, and so eDNA presents a rapid and cost-effective tool for applied conservation biology (Minamoto *et al*. 2012; Yoccoz *et al*, 2012a;

Barnes and Turner, 2016) with potential to be implemented in Southeast Asia. A recent annual horizon scan of global conservation issues (Sutherland *et al*. 2013) identified eDNA as one of the fifteen key topics that may increasingly impact upon conservation of biological diversity. There have been many studies applying eDNA to the detection of species of conservation concern some of which are mentioned below in Table 4, although there has not been a significant number of studies that have attempted to employ eDNA directly for management decisions.

Species	Assessment	Reference
Eastern Hellbender (*Cryptobranchus alleganiensis*)	Near threatened	Olson *et al*. 2012
Long-Finned Pilot Whale (*Globicephala melas*)	Data deficient	Foote *et al*. 2012
Bull Trout (*Salvelinus confluentus*)	Endangered	Wilcox *et al*. 2013, and 2014
Great Crested Newt (*Triturus cristatus*)	Least concern, but highly protected	Biggs *et al*. 2014, Rees *et al*. 2014
European Weather Loach (*Misgurnus fossilis*)	Least concern, but described as near-extinct in study paper	Sigsgaard *et al*. 2015
Chinook Salmon (*Oncorhynchus tshawytscha*)	Endangered	Laramie *et al*. 2015
Largetooth Sawfish (*Pristis microdon*)	Critically endangered	Simpfendorfer *et al*. 2016
Freshwater Pearl Mussel (*Margaritifera margaritifera*)	Endangered	Stoeckle *et al*. 2016
Pacific Lamprey (*Entosphenus tridentatus*)	Believed to be extinct in the wild	Carim *et al*. 2017
Chilean Devil Ray (*Mobula tarapacana*)	Vulnerable	Gargan *et al*. 2017
Aquatic heteropteran insect, *Nepa hoffmanni*	Endangered	Doi *et al*, 2017a
Maugean Skate (*Zearaja maugeana*)	Endangered	Weltz *et al*. 2017
Yangtze Finless Porpoise (*Neophocaena asiaeorientalis asiaeorientalis*)	Critically Endangered	Qu and Stewart, 2017, Stewart *et al*. 2017
Olm *(Proteus anguinus*)	Vulnerable	Vörös *et al*. 2017
West Indian Manatee *(Trichechus manatus)*, Amazonian Manatee *(Trichechus inunguis)*, West African Manatee *Trichechus senegalensis,*	Vulnerable	Hunter *et al*. 2018
Bull Shark (*Carcharhinus leucas*), Silly Shark (*C. falciformis*), Hardnose Shark (*C. macloti*), Spottail Shark (*C. sorrah*), Copper Shark (*C. brachyurus*) etc…	Vulnerable and Near Threatened	Boussarie-Bakker *et al*. 2018

Table 1.4. Examples of some eDNA studies detecting species of conservation concern. Assessment from the IUCN, (2017).

The next step is to go further than mere detection, and make conservation recommendations based on eDNA information. Pfleger *et al*. (2016) have recently done so, for example, after

successfully detecting the critically endangered Alabama Sturgeon (*Scaphirhynchus suttkusi*) and near threatened Gulf Sturgeon (*Acipenser oxyrinchus desotoi*) in the Mobile River Basin of Alabama, USA, using eDNA. They found that the distribution and temporal data suggested that both species migrated past navigation locks or dams, and remained upstream of passage barriers. The authors recommended that the removal of the barriers to passage would aid in the conservation of these species.

Some ichthyologists have defined the Southeast Asian/Eastern China region as the 'centre of dispersal' of the world's freshwater fishes (Wang *et al*. 1981, Menon, 1987). The Nagao Natural Environment Foundation's 'Fishes of Mainland Southeast Asia' (Kano *et al*. 2013) lists 757 defined species within 93 families within mainland Southeast Asia alone. However, Kottelat (2013) states that there are now 3,108 valid and named species within 137 families living in the inland waters of Southeast Asia, a figure that Kottelat predicts to only increase further as survey efforts increase and technologies improve. Some of the most charismatic aquatic species of conservation concern in Southeast Asia include Jullien's Golden Carp (*Probarbus jullieni*), the Narrow Saw-Fish (*Anoxypristis cuspidate*), the Mekong Giant Catfish (*Pangasianodon gigas*), Cantor's Giant Softshell Turtle (*Pelochelys cantorii*), the Giant Freshwater Stingray (*Himantura chaophraya*), and the False Gharial (*Tomistoma schlegelii*). Myers (2000) lists 568 species of amphibian; 750 species of reptile; and 422 species of mammals within the 'hotspots' of Sundaland, Wallacea, the Philippines and Indo-Burma, indicating that Southeast Asia is indeed exceptionally species rich. Information on the distribution of these numerous rare or endangered species, particularly for providing evidence to protect their associated habitats in these 'biodiversity hotspots' and propose conservation applications, is essential yet challenging (Lodge *et al*. 2012). eDNA and metabarcoding may provide an avenue for achieving this ambitious goal (Ji *et al*. 2013).

Non-invasive samples collected directly from e.g. faeces, egg shells, feathers and hair, although not eDNA *per se* (see Figure 1.2), have been used for population genetic analysis for some time (Beja-Pereira *et al*. 2009). The use of sloughed skin from the Humpback Whale, Sperm Whale and the North Atlantic Right Wale (*Megaptera novaeangliae, Physeter macrocephalus* and *Eubalaena glacialis*) by Amos *et al*. (1992) was a step towards true aquatic eDNA sampling, and provided population genetic data for conservation purposes. Building on the back of such population genetics studies from non-invasive samples, the use of eDNA in population genetics has very recently been achieved by Kapoor *et al*. (2017) and Afshinnekoo *et al*. (2015) to analyse human population diversity, and Sigsgaard *et al*. (2016)

to analyse whale shark population variation. Combined with the ability recently demonstrated by Deiner *et al*. (2017) to sequence entire mitogenomes of vertebrate eDNA, the applications of eDNA in conservation genetics and phylogeography is now as broad as current genomics techniques allow. This provides opportunities for estimating population size, population genetic relationships, species hybrids, and evolutionary patterns in samples of mixed genetic material (Barnes and Turner, 2016, Coissac *et al.* 2016), although discriminating between closely-related individuals from the same population will likely remain challenging in the near future.

Conservation efforts using eDNA may maximise success by incorporating data on temporal changes e.g. mating or die-offs (Barnes and Turner, 2016) by repeated sampling over time. Some very recent studies have successfully done so, demonstrating accurate seasonal fluctuations in e.g. newt eDNA concentration (Buxton *et al*. 2017b), invertebrate biodiversity (Bista *et al*., 2017), local migrations of native and non-native carp (Uschii *et al*. 2017) and jellyfish (Japanese Sea Nettle *Chrysaora pacifica*) presence (Minamoto *et al*. 2017) amongst others (Goldberg *et al*. 2011; Vervoort *et al*. 2012; de Souza *et al*. 2016; Sigsgaard *et al*. 2017; Stoeckle *et al*. 2017; Wu *et al*. 2018). Spatial changes in the detection of biodiversity using eDNA have also been observed, such as the change in local distribution of the Yangtze Finless Porpoise (*Neophocaena asiaeorientalis asiaeorientalis*), which was restricted to a core area of the Tian e-Zhou National Nature Reserve in Hubei, China during the breeding season (spring), but post-breeding eDNA concentrations were widespread across the reserve, encompassing sites previously thought to be unfrequented by the species (Stewart *et al*. 2017). This type of eDNA information may be linked to understanding of ecological processes which impact conservation such as habitat connectivity for migrating fish, recently explored by Yamanaka *et al*. (2016b), or on a fine scale, habitat use over particular life history events such as fish spawning.

When there is *a priori* knowledge of a habitat preference or behavioural pattern of the desired species, targeted sampling of specific microhabitats can allow eDNA detection of rare species in SEA. For example, eDNA detection of the golden tree frog from bromeliad water in Trinidad (Torresdal *et al*. 2017) would be a highly transferable approach to detect amphibians in the rainforests of Southeast Asia, such as Borneo's recently described Matang Narrow-Mouthed Frog (*Microhyla nepenthicola*) (Das and Haas, 2010), an obligate of the Pitcher Plant (*Nepenthes ampullaria*). The ability to detect mammals from leeches (Schnell *et al.* 2010), saliva left on browsed twigs (Nichols *et al.* 2012), and salt licks (Ishige *et al.* 2017)

would also be highly applicable to monitoring biodiversity in SEA where elusive mammals such as the Bornean Orangutan (*Pongo pygmaeus*), Sunda Pangolin (*Manis javanica*), Bornean Banteng (*Bos javanicus lowi*), and Saola (*Pseudoryx nghetinhensis*) are difficult to detect. With the ability of NGS technology to combine many samples, an obvious solution for conservation biologists with limited funding would be to maximise sampling of biodiversity by combining the targeting of multiple habitat types *and* specific microhabitats such as these all at once to monitor biodiversity of entire ecosystems, rather than focusing on individual species, taxon groups, or particular habitats; the approach adopted to date. Managers, agencies and researchers should have strong incentives to adopt eDNA monitoring techniques in conservation management as it provides rapid, cost-effective and reliable data with no *a priori* selection of target organisms. These techniques offer the opportunity to inform on and implement laws and regulations concerning management of natural resources, such as the establishment of a protected species e.g. the Great Crested Newt which triggered a suite of protection activity (Kelly *et al*. 2014a, Barnes and Turner 2016).

1.17 Conclusion

Although eDNA may be a novel, sensitive, species-specific and cost-effective tool with the potential to radically improve the detection of biodiversity, as discussed here, there is still much work to be done to improve this methodology to a level that may be reliably used in wildlife management. Currently, the field of eDNA is in the developmental stage (Dejean *et al*. 2012), with remaining gaps in the knowledge of how field and laboratory protocols influence the detection of eDNA, as well as how environmental conditions affect the production, degradation and detection of eDNA (Lodge *et al*. 2012). From a management perspective, levels of uncertainty that currently exist must be understood and communicated, especially when eDNA methodology is being used to inform management decisions, which can result in controversy, extreme scrutiny and in some cases, may even present legal challenges (Darling *et al*. 2011). The responsibility for participation in this communication falls with the stake-holders, method developers, resource managers, policy makers and public users of the specific ecosystem services (e.g. aquatic resources), who must engage in a transparent and informed discussion of the advantages and disadvantages of the use of eDNA in management decisions (Darling and Mahon. 2011), which will hopefully, after further experimental studies, be fully realized.

1.18 References

Abell *et al*. 2008. Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. *Bioscience*.

Afshinnekoo, E., Meydan, C., Chowdhury, S. *et al* 2015. Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Cell Systems* 1:72–87

Agersnap, S. *et al.*, 2017. Monitoring of noble, signal and narrow-clawed crayfish using environmental DNA from freshwater samples H. Doi, ed. *PLoS ONE*, 12(6), p.e0179261.

Alberdi, A., Aizpurua, O., Gilbert, M.T.P. and Bohmann, K., 2017. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. Journal of molecular biology, 215(3), pp.403-410.

Amano and Sutherland (2013). Four barriers to the global understanding of conservation: wealth, language, geographical location and security. *Proceedings of the Royal Society B*.

Amberg, J.J. *et al.*, 2015. Improving efficiency and reliability of environmental DNA analysis for silver carp. *Journal of Great Lakes Research*, 41(2).

Amos, W. *et al.*, 1992. Restrictable DNA From Sloughed Cetacean Skin; Its Potential for Use in Population Analysis. *Marine Mammal Science*, 8 (July), pp.275–283.

Andersen, K., Bird, K.L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjaer, K.H., Orlando, L., Gilbert, M.T.P. and Willerslev, E., 2012. Meta-barcoding of 'dirt'DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, *21*(8), pp.1966-1979.

Anderson-Carpenter, L.L., McLachlan, J.S., Jackson, S.T., Kuch, M., Lumibao, C.Y. and Poinar, H.N., 2011. Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. *BMC evolutionary biology*, *11*(1), p.30.

Andruszkiewicz, E.A., Starks, H.A., Chavez, F.P., Sassoubre, L.M., Block, B.A. and Boehm, A.B., 2017. Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. *PLoS ONE*, 12(4), p.e0176343.

Andújar, C., Arribas, P., Yu, D.W., Vogler, A.P. and Emerson, B.C., 2018. Why the COI barcode should be the community DNA metabarcode for the Metazoa. Molecular ecology.

Ardura, A. *et al.*, 2015a. eDNA and specific primers for early detection of invasive species - A case study on the bivalve *Rangia cuneata*, currently spreading in Europe. *Marine Environmental Research*, 112.

Ardura, A. *et al.*, 2015b. Environmental DNA evidence of transfer of North Sea molluscs across tropical waters through ballast water. *Journal of Molluscan Studies*, 81(4).

Asian Development Bank. 2016. *Indonesia: Country water assessment*. Mandaluyong City, Philippines: Asian Development Bank, 2016

Aylagas, E. *et al.*, 2016. Benchmarking DNA Metabarcoding for Biodiversity-Based Monitoring and Assessment. *Frontiers in Marine Science*, 3 (November).

Aylagas, E., Borja, Á. & Rodríguez-Ezpeleta, N., 2014. Environmental status assessment using DNA metabarcoding: Towards a genetics based marine biotic index (gAMBI*). PLoS ONE*, 9(3).

Baird, D.J. & Hajibabaei, M., 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Molecular Ecology, 21(8), pp.2039–2044.

Baker, C.S., Steel, D., Nieukirk, S. and Klinck, H., 2018. Environmental DNA (eDNA) From the Wake of the Whales: Droplet Digital PCR for Detection and Species Identification. Frontiers in Marine Science, 5, p.133.

Bakker, J., Wangensteen, O.S., Chapman, D.D., Boussarie, G., Buddo, D., Guttridge, T.L., Hertler, H., Mouillot, D., Vigliola, L. and Mariani, S., 2017. Environmental DNA reveals tropical shark diversity in contrasting levels of anthropogenic impact. *Scientific reports*, 7(1), p.16886.

Balian, *et al.* 2008. The Freshwater Animal Diversity Assessment: an overview of the results. *Hydrobiologia*. 595:627–637

Bálint, M. *et al.*, 2017. Twenty-five species of frogs in a liter of water: eDNA survey for exploring tropical frog diversity. , pp.0–36.

Barberán, A., Fernándz-Guerra, A., Bohannan, B.J. and Casamayor, E.O., 2012. Exploration of community traits as ecological markers in microbial metagenomes. *Molecular ecology*, *21*(8), pp.1909-1917.

Barnes, M.A. & Turner, C.R., 2016. The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, 17(1), pp.1–17.

Barnes, M.A., Turner, C.R., Jerde, C.L. *et al.* 2014. Environmental conditions influence eDNA persistence in aquatic systems. Environ Science and Technology 48:1819–1827.

Bass, D. *et al.*, 2015. Diverse Applications of Environmental DNA Methods in Parasitology. *Trends in Parasitology*, 31(10).

Beebee, T.J.C., 1991. Analysis, purification and quantification of extracelluIar DNA from aquatic environments. *Freshwater Biology*.

Beja-Pereira, A., Oliveira, R., Alves, P.C., *et al.* 2009. Advancing ecological understandings through technological transformations in noninvasive genetics. Mol Ecol Resour 9:1279–1301.

Berry, T.E., Osterrieder, S.K., Murray, D.C., Coghlan, M.L., Richardson, A.J., Grealy, A.K., Stat, M., Bejder, L. and Bunce, M., 2017. DNA metabarcoding for diet analysis and biodiversity: A case study using the endangered Australian sea lion (*Neophoca cinerea*). *Ecology and evolution*, 7(14), pp.5435-5453.

Bienert, F., De Danieli, S., Miquel, C. *et al.* 2012. Tracking earthworm communities from soil DNA. Molecular Ecology 21:2017–2030.

Biggs, J. *et al.*, 2015. Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). *Biological Conservation*, 183.

Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Griffiths, R.A., Foster, J., Wilkinson, J., Arnett, A., Williams, P. and Dunn, F., 2014. Analytical and methodological development for improved surveillance of the Great Crested Newt. *Defra Project WC1067. Freshwater Habitats Trust:* Oxford.

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

Binladen, J., Gilbert, M.T.P., Bollback, J.P., Panitz, F., Bendixen, C., Nielsen, R. and Willerslev, E., 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE*, *2*(2), p.e197.

Bista, I. *et al.*, 2017. Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, 8.

Bista, I., Carvalho, G.R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., Shokralla, S., Seymour, M., Bradley, D., Liu, S. and Christmas, M., 2018. Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. Molecular ecology resources.

Blaalid, R., Carlsen, T.O.R., Kumar, S., Halvorsen, R., Ugland, K.I., Fontana, G. and KAUSERUD, H., 2012. Changes in the root-associated fungal communities along a primary succession gradient analysed by 454 pyrosequencing. *Molecular ecology*, *21*(8), pp.1897-1908.

Bohan, D.A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A.J. and Woodward, G., 2017. Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks. *Trends in Ecology & Evolution*.

Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Douglas, W.Y. and De Bruyn, M., 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, 29(6), pp.358–367.

Bohmann, K., Monadjem, A., Noer, C.L., Rasmussen, M., Zeale, M.R., Clare, E., Jones, G., Willerslev, E. and Gilbert, M.T.P., 2011. Molecular diet analysis of two African free-tailed bats (Molossidae) using high throughput sequencing. *PLoS ONE*, *6*(6), p.e21441.

Borrell, Y.J., Miralles, L., Mártinez-Marqués, A., Semeraro, A., Arias, A., Carleos, C.E. and García-Vázquez, E., 2017. Metabarcoding and post-sampling strategies to discover non-indigenous species: A case study in the estuaries of the central south Bay of Biscay. *Journal for Nature Conservation*.

Boussarie, G., Bakker, J., Wangensteen, O.S., Mariani, S., Bonnin, L., Juhel, J.B., Kiszka, J.J., Kulbicki, M., Manel, S., Robbins, W.D. and Vigliola, L., 2018. Environmental DNA illuminates the dark diversity of sharks. *Science advances*, 4(5), p.eaap9661.

Bronnenhuber, J.E. & Wilson, C.C., 2013. Combining species-specific COI primers with environmental DNA analysis for targeted detection of rare freshwater species. Conservation Genetics Resources, 5(4).

Brook, B.W., Sodhi, N.S. and Ng, P.K., 2003. Catastrophic extinctions follow deforestation in Singapore. *Nature*, 424(6947), p.420.

Buxton, A.S. *et al.*, 2017a. Is the detection of aquatic environmental DNA influenced by substrate type? H. Doi, ed. *PLoS ONE*, 12(8), p.e0183371.

Buxton, A.S. *et al.*, 2017b. Seasonal variation in environmental DNA in relation to population size and environmental factors. *Scientific Reports*, 7, p.46294.

Buxton, A.S., Groombridge, J.J. and Griffiths, R.A., 2018. Seasonal variation in environmental DNA detection in sediment and water samples. *PloS one*, 13(1), p.e0191737.

Bylemans, J. *et al.*, 2016. An environmental DNA (eDNA) based method for monitoring spawning activity: a case study using the endangered Macquarie perch (*Macquaria australasica*). *Methods in Ecology and Evolution*.

Cai, W. *et al.*, 2017. Using eDNA to detect the distribution and density of invasive crayfish in the HongheHani rice terrace World Heritage site. *PLoS ONE*, 12(5).

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. and Holmes, S.P., 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nature methods, 13(7), p.581.

Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kühl, H., Boesch, C., Kappeler, P.M., Metzger, S., Schubert, G. and Leendertz, F.H., 2013. Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Molecular ecology*, *22*(4), pp.915-924.

Cannon, M.V., Hester, J., Shalkhauser, A., Chan, E.R., Logue, K., Small, S.T. and Serre, D., 2016. In silico assessment of primers for eDNA studies using PrimerTree and application to characterize the biodiversity surrounding the Cuyahoga River. *Scientific reports*, 6, p.22908.

Carim, K.J. *et al.*, 2017. A noninvasive tool to assess the distribution of Pacific lamprey (Entosphenus tridentatus) in the Columbia River basin. *PLoS ONE*, 12(1).

Chariton, A.A., Court, L.N., Hartley, D.M., Colloff, M.J., Hardy, C.M. 2010. Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Frontiers in Ecology and the Environment*, 8, 233–238.

Chariton, A.A., Ho, K.T., Proestou, D., Bik, H., Simpson, S.L., Portis, L.M., Cantwell, M.G., Baguley, J.G., Burgess, R.M., Pelletier, M.M. and Perron, M., 2014. A molecular-based approach for examining responses of eukaryotes in microcosms to contaminant-spiked estuarine sediments. *Environmental Toxicology and Chemistry*, *33*(2), pp.359-369.

Chen, W., Zhang, C.K., Cheng, Y., Zhang, S. and Zhao, H., 2013. A comparison of methods for clustering 16S rRNA sequences into OTUs. *PloS one*, *8*(8), p.e70837.

Civade, R., Dejean, T., Valentini, A., Roset, N., Raymond, J.C., Bonin, A., Taberlet, P. and Pont, D., 2016. Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. *PloS one*, 11(6), p.e0157366.

Clare, E.L., Fraser, E.E., Braid, H.E., Fenton, M.B. and Hebert, P.D., 2009. Species on the menu of a generalist predator, the eastern red bat (Lasiurus borealis): using a molecular approach to detect arthropod prey. *Molecular ecology*, *18*(11), pp.2532-2542.

Clarke, P. and Jupiter, S., 2010. *Principles and practice of ecosystem-based management: a guide for conservation practitioners in the tropical western pacific*. Wildlife Conservation Society.

Clarke, L.J., Beard, J.M., Swadling, K.M. and Deagle, B.E., 2017. Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. *Ecology and evolution*, 7(3), pp.873-883.

Coissac, E. *et al.*, 2016. From barcodes to genomes: Extending the concept of DNA barcoding. *Molecular Ecology*, 25(7).

Collen, B., Whitton, F., Dyer, E.E., Baillie, J.E., Cumberlidge, N., Darwall, W.R., Pollock, C., Richman, N.I., Soulsby, A.M. and Böhm, M., 2014. Global patterns of freshwater species diversity, threat and endemism. *Global Ecology and Biogeography*, 23(1), pp.40-51.

Collins, R.A. *et al.*, 2013. Something in the water: Biosecurity monitoring of ornamental fish imports using environmental DNA. *Biological Invasions,* 15(6).

Collins, R. A., Wangensteen, O. S., O'Gorman, E. J., Mariani, S., Sims, D. W., & Genner, M. J. 2018. Persistence of environmental DNA in marine systems. *Communications Biology*. Volume 1, Article number: 185.

Corinaldesi, C., Barucca, M., Luna, G.M. and DELL'ANNO, A., 2011. Preservation, origin and genetic imprint of extracellular DNA in permanently anoxic deep-sea sediments. *Molecular ecology*, *20*(3), pp.642-654.

Costanza, R., d'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'neill, R.V., Paruelo, J. and Raskin, R.G., 1997. The value of the world's ecosystem services and natural capital. nature, 387(6630), pp.253-260.

Creer, S. *et al.*, 2016. The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution.*

Creer, S., Fonseca, V.G., Porazinska, D.L., GIBLIN-DAVIS, R.M., Sung, W., Power, D.M., Packer, M., Carvalho, G.R., Blaxter, M.L., Lambshead, P.J.D. and Thomas, W.K., 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology*, 19(s1), pp.4-20.

Cristescu, M.E. and Hebert, P.D., 2018. Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation. *Annual Review of Ecology, Evolution, and Systematics*, (0).

Cristescu, M.E., 2014. From barcoding single individuals to metabarcoding biological communities: Towards an integrative approach to the study of global biodiversity. *Trends in Ecology and Evolution*, 29(10).

Cumberlidge, N., Ng, P.K., Yeo, D.C., Magalhães, C., Campos, M.R., Alvarez, F., Naruse, T., Daniels, S.R., Esser, L.J., Attipoe, F.Y. and Clotilde-Ba, F.L., 2009. Freshwater crabs and the biodiversity crisis: importance, threats, status, and conservation challenges. *Biological Conservation*, 142(8), pp.1665-1673.

Dalén, L., Götherström, A., Meijer, T. and Shapiro, B., 2007. Recovery of DNA from footprints in the snow. *The Canadian Field-Naturalist*, *121*(3), pp.321-324.

Darling, J.A. and Mahon, A.R., 2011. From molecules to management: adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environmental research*, *111*(7), pp.978-988.

Darling, J.A., Galil, B.S., Carvalho, G.R., Rius, M., Viard, F. and Piraino, S., 2017. Recommendations for developing and applying genetic tools to assess and manage biological invasions in marine ecosystems. Marine policy, 85, pp.54-64.

Das, I. and Haas, A., 2010. New species of Microhyla from Sarawak: Old World's smallest frogs crawl out of miniature pitcher plants on Borneo (Amphibia: Anura: Microhylidae). Zootaxa, 2571, pp.37-52.

Davison, P.I. *et al.*, 2017. Application of environmental DNA analysis to inform invasive fish eradication operations. *The Science of Nature*, 104(3–4), p.35.

De Barba, M. *et al.*, 2014. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular Ecology Resources*, 14(2), pp.306–323.

De Bruyn, M., Stelbrink, B., Morley, R.J., Hall, R., Carvalho, G.R., Cannon, C.H., Van Den Bergh, G., Meijaard, E., Metcalfe, I., Boitani, L. and Maiorano, L. 2014. Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity. *Systematic Biology*, 63(6), pp.879-901.

De Souza, L.S. *et al.*, 2016. Environmental DNA (eDNA) Detection Probability Is Influenced by Seasonal Activity of Organisms H. Doi, ed. *PLoS ONE*, 11(10), p.e0165273.

De Ventura, L. *et al.*, 2017. Tracing the quagga mussel invasion along the Rhine river system using eDNA markers: early detection and surveillance of invasive zebra and quagga mussels. *Management of Biological Invasions*, 8(1).

Deagle, B.E., Kirkwood, R. and Jarman, S.N., 2009. Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular ecology*, *18*(9), pp.2022-2038.

Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F. & Taberlet, P. 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, 10, 20140562.

Deck, J., Gaither, M.R., Ewing, R., Bird, C.E., Davies, N., Meyer, C., Riginos, C., Toonen, R.J. and Crandall, E.D., 2017. The Genomic Observatories Metadatabase (GeOMe): A new repository for field and sampling event metadata associated with genetic samples. PLoS Biology, 15(8), p.e2002925.

Deiner, K. and Altermatt, F., 2014. Transport distance of invertebrate environmental DNA in a natural river. *PLoS ONE*, *9*(2), p.e88786.

Deiner, K., Walser. J., 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. and Altermatt, F., 2016. Environmental DNA reveals that rivers are conveyer belts of biodiversity information. *Nature communications*, 7.

Deiner, K., Renshaw, M.A., Li, Y., Olds, B.P., Lodge, D.M. and Pfrender, M.E., 2017a. Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. *Methods in Ecology and Evolution*.

Deiner, K., *et al.* 2017b. Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Molecular Ecology*.

Dejean, T., Valentini, A., Duparc, A., Pellier-Cuit, S., Pompanon, F., Taberlet, P. and Miaud, C., 2011. Persistence of environmental DNA in freshwater ecosystems. *PLoS ONE*, *6*(8), p.e23398.

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E. and 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*(4), pp.953-959.

Djurhuus, A., Port, J., Closek, C.J., Yamahara, K.M., Romero-Maraccini, O., Walz, K.R., Goldsmith, D.B., Michisaki, R., Breitbart, M., Boehm, A.B. and Chavez, F.P., 2017. Evaluation of Filtration and DNA Extraction Methods for Environmental DNA Biodiversity Assessments across Multiple Trophic Levels. *Frontiers in Marine Science*, 4, p.314.

Doi, H. *et al.*, 2017a. Detection of an endangered aquatic heteropteran using environmental DNA in a wetland ecosystem. *Royal Society Open Science*, 4(7).

Doi, H. *et al.*, 2017b. Environmental DNA analysis for estimating the abundance and biomass of stream fish. *Freshwater Biology*, 62(1).

Doi, H., Akamatsu, Y., Watanabe, Y., Goto, M., Inui, R., Katano, I., Nagano, M., Takahara, T. and Minamoto, T. 2017c. Water sampling for environmental DNA surveys by using an unmanned aerial vehicle. *Limnology and Oceanography: Methods*.

Dougherty, M.M. *et al.*, 2016. Environmental DNA (eDNA) detects the invasive rusty crayfish *Orconectes rusticus* at low abundances. *Journal of Applied Ecology*, 53(3).

Drinkwater, R., Schnell, I.B., Bohmann, K., Bernard, H., Veron, G., Clare, E., Gilbert, M.T.P. and Rossiter, S.J., 2018. Using metabarcoding to compare the suitability of two blood-feeding leech species for sampling mammalian diversity in North Borneo. *Molecular Ecology Resources.*

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. and Sullivan, C.A., 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews*, 81(2), pp.163-182.

Dunker, K.J. *et al.*, 2016. Potential of environmental DNA to evaluate northern pike (*Esox lucius*) eradication efforts: An experimental test and case study. *PLoS ONE*, 11(9).

Dunn, N., Priestley, V., Herraiz, A., Arnold, R. and Savolainen, V., 2017. Behavior and season affect crayfish detection and density inference using environmental DNA. *Ecology and Evolution*.

Egan, S.P. *et al.*, 2015. Rapid molecular detection of invasive species in ballast and harbour water by integrating environmental DNA and light transmission spectroscopy. *Environmental Science and Technology*, 49(7).

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 and Technology*, 50(4).

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(1), pp.56–68.

Eiler, A., Löfgren, A., Hjerne, O., Nordén, S. and Saetre, P., 2018. Environmental DNA (eDNA) detects the pool frog (Pelophylax lessonae) at times when traditional monitoring methods are insensitive. *Scientific Reports*, 8(1), p.5452.

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(7).

Elbrecht, V. and Leese, F., 2017. PrimerMiner: an R package for development and in silico validation of DNA metabarcoding primers. *Methods in Ecology and Evolution*, 8(5), pp.622-626.

Epp, L.S., Boessenkool, S., Bellemain, E.P., Haile, J., Esposito, A., Riaz, T., Erseus, C., Gusarov, V.I., Edwards, M.E., Johnsen, A. and Stenøien, H.K., 2012. New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular ecology*, *21*(8), pp.1821-1833.

Erickson, R.A. *et al.*, 2016. Detecting the movement and spawning activity of bigheaded carps with environmental DNA. *Molecular Ecology Resources*, 16(4).

Esposito, A., Kirschberg, M. 2014. How many 16S-based studies should be included in a metagenomic conference? It may be a matter of etymology. *FEMS Microbiology Letters* 351:145–146.

Evans, N.T. *et al.*, 2016. Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*, 16(1), pp.29–41.

Evans, N.T. and Lamberti, G.A., 2017. Freshwater fisheries assessment using environmental DNA: A primer on the method, its potential, and shortcomings as a conservation tool. *Fisheries Research*.

Evans, N.T. *et al.*, 2017a. Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, p.cjfas-2016-0306.

Evans, N.T. *et al.*, 2017b. Comparative Cost and Effort of Fish Distribution Detection via Environmental DNA Analysis and Electrofishing. *Fisheries*, 42(2), pp.90–99.

Everett, M. V., Park, L. K. 2017. Exploring deep-water coral communities using environmental DNA. *Deep Sea Research Part II: Topical Studies in Oceanography*.

Fahner, N.A. *et al.*, 2016. Large-scale monitoring of plants through environmental DNA metabarcoding of soil: Recovery, resolution, and annotation of four DNA markers. *PLoS ONE*, 11(6).

Fahner, N.A., McCarthy, A., Barnes, J.G., Singer, G. and Hajibabaei, M., 2018. Experimental design considerations for assessing marine biodiversity using environmental DNA (No. e26814v1). *PeerJ Preprints.*

Faust, K. and Raes, J., 2012. Microbial interactions: from networks to models. *Nature reviews. Microbiology*, *10*(8), p.538.

Ficetola, G.F., Miaud, C., Pompanon, F. and Taberlet, P., 2008. Species detection using environmental DNA from water samples. *Biology letters*, *4*(4), pp.423-425.

Ficetola, G.F. *et al.*, 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, 15(3), pp.543–556.

Ficetola, G.F., Taberlet, P. & Coissac, E., 2016. How to limit false positives in environmental DNA and metabarcoding? *Molecular Ecology Resources*, 16(3).

Floyd, R., Abebe, E., Papert, A., Blaxter, M. 2002. Molecular barcodes for soil nematode identification. *Molecular Ecology* 11: 839–850.

Folloni, S., Kagkli, D., Rajcevic, B., Guimaraes, N.C., Van Droogenbroeck, B., Valicente, F.H., Van den Eede, G.U.Y. and Van den Bulcke, M., 2012. Detection of airborne genetically modified maize pollen by real-time PCR. *Molecular ecology resources*, *12*(5), pp.810-821.

Fonseca, V.G., Carvalho, G.R., Sung, W., Johnson, H.F., Power, D.M., Neill, S.P., Packer, M., Blaxter, M.L., Lambshead, P.J.D., Thomas, W.K., Creer, S. 2010. Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications*

Foote, A.D., Thomsen, P.F., Sveegaard, S., Wahlberg, M., Kielgast, J., Kyhn, L.A., Salling, A.B., Galatius, A., Orlando, L. and Gilbert, M.T.P., 2012. Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. *PLoS ONE*, *7*(8), p.e41781.

Frøslev, T.G., Kjøller, R., Bruun, H.H., Ejrnæs, R., Brunbjerg, A.K., Pietroni, C. and Hansen, A.J., 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature communications*, 8(1), p.1188.

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(2).

Furlan, E.M. & Gleeson, D., 2017. Improving reliability in environmental DNA detection surveys through enhanced quality control. *Marine and Freshwater Research*, 68(2), p.388.

Gansauge, M.T. and Meyer, M., 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature protocols*, *8*(4), p.737.

Gantz, C.A., Renshaw, M.A., Erickson, D., Lodge, D.M. and Egan, S.P., Environmental DNA detection of aquatic invasive plants in lab mesocosm and natural field conditions. *Biological Invasions*, pp.1-18.

Gargan, L.M. *et al.*, 2017. Development of a sensitive detection method to survey pelagic biodiversity using eDNA and quantitative PCR: a case study of devil ray at seamounts. *Marine Biology*, 164(5).

Gariepy, T.D., Lindsay, R., Ogden, N. and Gregory, T.R., 2012. Identifying the last supper: utility of the DNA barcode library for bloodmeal identification in ticks. *Molecular ecology resources*, *12*(4), pp.646-652.

Ghosal, R., Eichmiller, J.J., Witthuhn, B.A. and Sorensen, P.W., 2018. Attracting Common Carp to a bait site with food reveals strong positive relationships between fish density, feeding activity,

environmental DNA, and sex pheromone release that could be used in invasive fish management. *Ecology and Evolution*.

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.J., Griggo, C., Gielly, L., Domaizon, I., Coissac, E., David, F., Choler, P. and Poulenard, J., 2014. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature communications*, *5*, p.3211.

Gilbert, M.T.P., Bandelt, H.J., Hofreiter, M. and Barnes, I., 2005. Assessing ancient DNA studies. *Trends in ecology & evolution*, *20*(10), pp.541-544.

Gingera, T.D. *et al.*, 2017. Environmental DNA as a detection tool for zebra mussels Dreissena polymorpha (Pallas, 1771) at the forefront of an invasion event in Lake Winnipeg, Manitoba, Canada. *Management of Biological Invasions*, 8

Goldberg, C.S., Pilliod, D.S., Arkle, R.S. and 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*(7), p.e22746.

Goldberg, C.S. *et al.*, 2013. Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshwater Science*, 32(3).

Goldberg, C.S. *et al.*, 2016. Critical considerations for the application of environmental DNA methods to detect aquatic species. M. Gilbert, ed. *Methods in Ecology and Evolution*, 7(11), pp.1299–1307.

Goldberg, C.S., Strickler, K.M. and Fremier, A.K., 2018. Degradation and dispersion limit environmental DNA detection of rare amphibians in wetlands: Increasing efficacy of sampling designs. *Science of The Total Environment*, 633, pp.695-703.

Gomes, G.B. *et al.*, 2017. Use of environmental DNA (eDNA) and water quality data to predict protozoan parasites outbreaks in fish farms.

Grey, E.K., Bernatchez, L., Cassey, P., Deiner, K., Deveney, M., Howland, K.L., Lacoursière-Roussel, A., Leong, S.C.Y., Li, Y., Olds, B. and Pfrender, M.E., 2018. Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Scientific reports*, 8(1), p.8843.

Guardiola, M. *et al.*, 2015. Deep-sea, deep-sequencing: Metabarcoding extracellular DNA from sediments of marine canyons. *PLoS ONE*, 10(10).

Guinée, J., Kleijn, R. and Henriksson, P., 2010. Environmental life cycle assessment of South-East Asian aquaculture systems for tilapia, pangasius, catfish, penaeid shrimp and macrobrachium prawns. Goal & Scope Definition Report-Final version. *Sustaining Ethical Aquaculture Trade (SEAT)* Deliverable Ref: D, 2.

Haegeman, B. and Loreau, M., 2011. A mathematical synthesis of niche and neutral theories in community ecology. *Journal of Theoretical Biology*, *269*(1), pp.150-165.

Haile, J., Holdaway, R., Oliver, K., Bunce, M., Gilbert, M.T.P., Nielsen, R., Munch, K., Ho, S.Y., Shapiro, B. and Willerslev, E., 2007. Ancient DNA chronology within sediment deposits: Are paleobiological reconstructions possible and is DNA leaching a factor?. *Molecular biology and evolution*, *24*(4), pp.982-989.

Hails, C. *et al.* 2008. Living Planet Report 2008. WWF.

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.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A. and Baird, D.J., 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE*, *6*(4), p.e17497.

Hall, E.M. *et al.*, 2016. Evaluating environmental DNA-based quantification of ranavirus infection in wood frog populations. *Molecular Ecology Resources*, 16(2).

Handelsman, J., 2009. Metagenetics: spending our inheritance on the future. *Microbial biotechnology*, 2(2), pp.138-139.

Hänfling, B. *et al.*, 2016. Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods.

Hansen, B.K., Bekkevold, D., Clausen, L.W. and Nielsen, E.E. 2018. The sceptical optimist: challenges and perspectives for the application of environmental DNA in marine fisheries. *Fish and Fisheries.*

Hao, X., Jiang, R. and Chen, T., 2011. Clustering 16S rRNA for OTU prediction: a method of unsupervised Bayesian clustering. *Bioinformatics*, *27*(5), pp.611-618.

Harper, L.R., Handley, L.L., Hahn, C., Boonham, N., Rees, H.C., Lewis, E., Adams, I.P., Brotherton, P., Phillips, S. and Hänfling, B., 2018. Understanding biodiversity at the pondscape using environmental DNA: a focus on great crested newts. *bioRxiv*, p.278309.

Hartikainen, H. *et al.*, 2016. Assessing myxozoan presence and diversity using environmental DNA. *International Journal for Parasitology*, 46(12).

Hartman, L.J., Coyne, S.R. & Norwood, D.A. (2005) Development of a novel internal positive control for Taqman® based assays. *Molecular and Cellular Probes*, 19, 51–59.

Hebert, P.D., Ratnasingham, S. and de Waard, J.R., 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(Suppl 1), pp.S96-S99.

Hebsgaard, M.B., Gilbert, M.T.P., Arneborg, J., Heyn, P., Allentoft, M.E., Bunce, M., Munch, K., Schweger, C. and Willerslev, E., 2009. 'The Farm Beneath the Sand'–an archaeological case study on ancient 'dirt' DNA. *Antiquity*, *83*(320), pp.430-444.

Helfman, G.S., 2007. *Fish conservation: a guide to understanding and restoring global aquatic* biodiversity and fishery resources. Island Press.

Hering, D., Borja, A., Jones, J.I., Pont, D., Boets, P., Bouchez, A., Bruce, K., Drakare, S., Hänfling, B., Kahlert, M. and Leese, F., 2018. Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. *Water research.*

Hiiesalu, I., Oepik, M., Metsis, M., Lilje, L. 2012. Plant species richness belowground: higher richness and new patterns revealed by next-generation sequencing. *Molecular Ecology* 21: 2004–2016.

Hinlo, R. *et al.*, 2017a. Environmental DNA monitoring and management of invasive fish: comparison of eDNA and fyke netting. *Management of Biological Invasions*, 8(1).

Hundermark, E.L. and Takahashi, M.K., 2018. Improving the yield of environmental DNA from filtered aquatic samples. Conservation Genetics Resources, pp.1-3.

Hunter, M.E. *et al.*, 2015. Environmental DNA (eDNA) sampling improves occurrence and detection estimates of invasive Burmese pythons. *PLoS ONE*, 10(4).

Hunter, M.E., Meigs-Friend, G., Ferrante, J.A., Kamla, A.T., Dorazio, R.M., Diagne, L.K., Luna, F., Lanyon, J.M. and Reid, J.P., 2018. Surveys of environmental DNA (eDNA): a new approach to estimate occurrence in Vulnerable manatee populations. *Endangered Species Research*, 35, pp.101-111.

Huver, J.R., Koprivnikar, J., Johnson, P.T.J. and Whyard, S., 2015. Development and application of an eDNA method to detect and quantify a pathogenic parasite in aquatic ecosystems. *Ecological Applications*, 25(4), pp.991-1002.

Ishige, T. *et al.*, 2017. Tropical-forest mammals as detected by environmental DNA at natural saltlicks in Borneo. *Biological Conservation.*

Ivanova, N.V., Zemlak, T.S., Hanner, R.H. and Hebert, P.D., 2007. Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Resources*, 7(4), pp.544-548.

Jane, S.F., Wilcox, T.M., McKelvey, K.S., Young, M.K., Schwartz, M.K., Lowe, W.H., *et al.* 2015. Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. Molecular Ecology Resources. 2015; 15(1):216– 227.

Jenkins, M., 2003. Prospects for biodiversity. *Science*, 302(5648), pp.1175-1177.

Jerde, C.L., Mahon, A.R., Chadderton, W.L. and Lodge, D.M., 2011. "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, 4(2), pp.150-157.

Jerde, C.L. *et al.*, 2016. Influence of Stream Bottom Substrate on Retention and Transport of Vertebrate Environmental DNA. *Environmental Science and Technology*, 50(16).

Ji, Y. *et al.*, 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding M. Holyoak, ed. *Ecology Letters*, 16(10), pp.1245–1257.

Jo, T. *et al.*, 2017. Rapid degradation of longer DNA fragments enables the improved estimation of distribution and biomass using environmental DNA. *Molecular Ecology Resources*.

Jones, M., Ghoorah, A. and Blaxter, M., 2011. jMOTU and taxonerator: turning DNA barcode sequences into annotated operational taxonomic units. *PLoS one*, *6*(4), p.e19259.

Jørgensen, T., Haile, J., Möller, P.E.R., Andreev, A., Boessenkool, S., Rasmussen, M., Kienast, F., Coissac, E., Taberlet, P., Brochmann, C. and Bigelow, N.H., 2012. A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular Ecology*, *21*(8), pp.1989-2003.

Jørgensen, T., Kjær, K.H., Haile, J., Rasmussen, M., Boessenkool, S., Andersen, K., Coissac, E., Taberlet, P., Brochmann, C., Orlando, L. and Gilbert, M.T.P., 2012. Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA Meta-barcoding. *Molecular Ecology*, *21*(8), pp.1980-1988.

Kano Y, *et al.* 2013. An online database on freshwater fish diversity and distribution in Mainland Southeast Asia. *Ichthyological Research* 60: 293-295.

Kapoor, V. *et al.*, 2017. Analysis of human mitochondrial DNA sequences from fecally polluted environmental waters as a tool to study population diversity. *AIMS Environmental Science*, 4(3), pp.443–455.

Katano, I. *et al.*, 2017. Environmental DNA method for estimating salamander distribution in headwater streams, and a comparison of water sampling methods M. Stöck, ed. *PLoS ONE*, 12(5), p.e0176541.

Kelly, R.P. *et al.*, 2014a. Harnessing DNA to improve environmental management. *Science*, 344(6191).

Kelly, R.P., Port, J.A., Yamahara, K.M. and Crowder, L.B., 2014b. Using environmental DNA to census marine fishes in a large mesocosm. *PLoS ONE*, *9*(1), p.e86175.

Kelly, R.P. *et al.*, 2016. Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ*, 4.

Kelly, R.P. *et al.*, 2017. Genetic and Manual Survey Methods Yield Different and Complementary Views of an Ecosystem. *Frontiers in Marine Science*, 3.

Kent, R.J., 2009. Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. *Molecular ecology resources*, *9*(1), pp.4-18.

Klymus, K.E. *et al.*, 2015. Quantification of eDNA shedding rates from invasive bighead carp Hypophthalmichthys nobilis and silver carp *Hypophthalmichthys molitrix*. *Biological Conservation,* 183.

Klymus, K.E., Marshall, N.T. and Stepien, C.A., 2017a. Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. *PloS ONE*, 12(5), p.e0177643.

Klymus, K.E., Richter, C.A., Thompson, N. and Hinck, J.E., 2017b. Metabarcoding of Environmental DNA Samples to Explore the Use of Uranium Mine Containment Ponds as a Water Source for Wildlife. *Diversity*, 9(4), p.54.

Knapp, M. and Hofreiter, M., 2010. Next generation sequencing of ancient DNA: requirements, strategies and perspectives. *Genes*, *1*(2), pp.227-243.

Knapp, M., Stiller, M. and Meyer, M., 2012. Generating barcoded libraries for multiplex high-throughput sequencing. *Ancient DNA: Methods and Protocols*, pp.155-170.

Kolar, C.S. and Lodge, D.M., 2001. Progress in invasion biology: predicting invaders. *Trends in ecology & evolution*, 16(4), pp.199-204.

Kottelat, M. and Whitten, T., 1996. Freshwater biodiversity in Asia: with special reference to fish (Vol. 343). *World Bank Publications*.

Kottelat, M. 2013. The fishes of the inland waters of Southeast Asia: a catalogue and core bibliography of the fishes known to occur in freshwaters, mangroves and estuaries. *National University of Singapore*.

La Salle, J., Williams, K.J. & Moritz, C., 2016. Biodiversity analysis in the digital era. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 371(1702).

Lacoursière-Roussel, A. *et al.*, 2016a. Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. *Journal of Applied Ecology*, 53(4).

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(6).

Lance, R.F. *et al.*, 2017. Experimental observations on the decay of environmental DNA from bighead and silver carps. *Management of Biological Invasions*, 8.

Laramie, M.B., Pilliod, D.S. & Goldberg, C.S., 2015. Characterizing the distribution of an endangered salmonid using environmental DNA analysis. *Biological Conservation*, 183.

Larson, E.R. *et al.*, 2017. Environmental DNA (eDNA) detects the invasive crayfishes Orconectes rusticus and Pacifastacus leniusculus in large lakes of North America. *Hydrobiologia*, pp.1–13.

Lawson Handley, L., 2015. How will the 'molecular revolution' contribute to biological recording?. Biological Journal of the Linnean Society, 115(3), pp.750-766.

Lefort, M.C., Boyer, S., Barun, A., Khoyi, A.E., Ridden, J., Smith, V.R., Sprague, R., Waterhouse, B.R. and Cruickshank, R.H., 2015. Blood, sweat and tears: non-invasive vs. non-disruptive DNA sampling for experimental biology (No. e1580). *PeerJ* PrePrints.

Lejzerowicz, F., Esling, P., Majewski, W., Szczuciński, W., Decelle, J., Obadia, C., Arbizu, P.M. and Pawlowski, J., 2013. Ancient DNA complements microfossil record in deep-sea subsurface sediments. *Biology letters*, *9*(4), p.20130283.

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

Leray, M. & Knowlton, N., 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, 2014 (July), p.201424997.

Leray, M., Knowlton, N. 2017. Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. *PeerJ* 5, e3006.

Lévêque, C., T. Oberdorff, D. Paugy, M.L.J. Stiassny & P.A. Tedesco. 2008. Global diversity of fish (Pisces) in freshwater. In: Balian E. V., C. Lévêque, H. Segers & K. Martens (eds), Freshwater Animal Diversity Assessment. *Hydrobiologia*

Li, F., Mahon, A.R., Barnes, M.A., *et al.* 2011. Quantitative and rapid DNA detection by laser transmission spectroscopy. *PLoS ONE*.

Li, J., Lawson Handley, L.J., Read, D.S. and Hänfling, B., 2018. The effect of filtration method on the efficiency of environmental DNA capture and quantification via metabarcoding. Molecular ecology resources.

Little, D.P. 2014. A DNA mini-barcode for land plants. *Molecular Ecology Resources* 14: 437–446.

Liu, S., Wang, X., Xie, L., Tan, M., Li, Z., Su, X., Zhang, H., Misof, B., Kjer, K.M., Tang, M. and Niehuis, O., 2016. Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. *Molecular Ecology Resources*, 16(2), pp.470-479.

Lodge, D.M., Williams, S., MacIsaac, H.J., Hayes, K.R., Leung, B., Reichard, S., Mack, R.N., Moyle, P.B., Smith, M., Andow, D.A. and Carlton, J.T., 2006. Biological invasions: recommendations for US policy and management. *Ecological applications*, 16(6), pp.2035-2054.

Lodge, D.M., Turner, C.R., Jerde, C.L., Barnes, M.A., Chadderton, L., Egan, S.P., Feder, J.L., Mahon, A.R. and Pfrender, M.E., 2012. Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology*, *21*(11), pp.2555-2558.

Lohman, D.J., de Bruyn, M., Page, T., von Rintelen, K., Hall, R., Ng, P.K., Shih, H.T., Carvalho, G.R. and von Rintelen, T., 2011. Biogeography of the Indo-Australian archipelago. *Annual*

Lugg, W.H., Griffiths, J., van Rooyen, A.R., Weeks, A.R. and Tingley, R., 2018. Optimal survey designs for environmental DNA sampling. *Methods in Ecology and Evolution*, 9(4), pp.1049-1059.

Mächler, 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. *et al.*, 2016. Fishing in the Water: Effect of Sampled Water Volume on Environmental DNA-Based Detection of Macroinvertebrates. *Environmental Science and Technology*, 50(1).

Madden, A.A., Barberán, A., Bertone, M.A., Menninger, H.L., Dunn, R.R. and Fierer, N., 2016. The diversity of arthropods in homes across the United States as determined by environmental DNA analyses. Molecular ecology.

Magurran, A.E., 2013. Measuring biological diversity. *John Wiley & Sons*.

Mahé, F., Rognes, T., Quince, C., de Vargas, C. and Dunthorn, M., 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ*, 3, p.e1420.

Mahon, A.R., Barnes, M.A., Li, F. *et al.* 2012. DNA-based species detection capabilities using laser transmission spectroscopy. *J R Soc Interface* 10:20120637.

Mahon, A.R., Jerde, C.L., Galaska, M., Bergner, J.L., Chadderton, W.L., Lodge, D.M., Hunter, M.E. and Nico, L.G., 2013. Validation of eDNA surveillance sensitivity for detection of Asian carps in controlled and field experiments. *PLoS ONE*, 8(3), p.e58316.

Mahon AR, Nathan LR, Jerde CL. 2014. Meta-genomic surveillance of invasive species in the bait trade. *Conservation Genetic Resources* 6:563–567.

Majaneva, M., Diserud, O.H., Eagle, S.H., Boström, E., Hajibabaei, M. and Ekrem, T., 2018. Environmental DNA filtration techniques affect recovered biodiversity. *Scientific reports*, 8(1), p.4682.

Manning, M.M. and Gates, R.D., 2008. Diversity in populations of free-living Symbiodinium from a Caribbean and Pacific reef. *Limnology and Oceanography*, 53(5), pp.1853-1861.

Martellini, A., Payment, P. and Villemur, R., 2005. Use of eukaryotic mitochondrial DNA to differentiate human, bovine, porcine and ovine sources in fecally contaminated surface water. *Water research*, 39(4), pp.541-548.

Matsuhashi, S. *et al.*, 2016. Evaluation of the environmental DNA method for estimating distribution and biomass of submerged aquatic plants. *PLoS ONE*, 11(6).

Matsui, K., Honjo, M., Kawabata, Z. 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

Menon, A.G.K. 1987. The fauna of India and the adjacent countries. Pisces, IV Teleostei-Cobitoidea, Part 1, Homalopteridae, x+259 pp, 16 plates. *Zoological Survey of India, Calcutta*.

Merkes, C.M. *et al.*, 2014. Persistence of DNA in carcasses, slime and avian faeces may affect interpretation of environmental DNA data. *PLoS ONE*, 9(11).

Metcalfe, I. 2011. Tectonic framework and Phanerozoic evolution of Sundaland. *Gondwana Research*. Vol. 19 (pg. 3-21)

Michelin, G., Heckly, X. and Rigaux, B., 2011. Rapport d'etude–ADN Environnemental, Detection de l'Espece Exotique Envahissante Grenouille Taureau. (Environmental Study - DNA Report, Invasive Alien Species Detection Bullfrog). *DREAL, CDPNE, CR Centre, Spygen*, p.20.

Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M.N. and Kawabata, Z.I., 2012. Surveillance of fish species composition using environmental DNA. *Limnology*, *13*(2), pp.193-197.

Minamoto, T. *et al.*, 2016. Techniques for the practical collection of environmental DNA: filter selection, preservation, and extraction. *Limnology*, 17(1).

Minamoto, T. *et al.*, 2017. Environmental DNA reflects spatial and temporal jellyfish distribution. *PLoS ONE*, 12(2), p.e0173073.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J.Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H. and Kondoh, M., 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society open science*, 2(7), p.150088.

Molik, D., Pfrender, M. and Emrich, S., 2018. The Effects of Normalization, Transformation, and Rarefaction on Clustering of OTU Abundance. *bioRxiv*, p.259325.

Morgan, M.J. *et al.*, 2013. Improved Inference of Taxonomic Richness from Environmental DNA. *PLoS ONE*, 8(8).

Morisset, D., Štebih, D., Milavec, M., Gruden, K. and Žel, J., 2013. Quantitative analysis of food and feed samples with droplet digital PCR. *PLoS ONE*, 8(5), p.e62583.

Morley, R.J. Gower, D., Johnson, K.G., Rosen, B.R., Richardson J., Rüber L., Williams, S.T. 2012. A review of the Cenozoic palaeoclimate history of Southeast Asia, Biotic evolution and environmental change in Southeast Asia. *Cambridge University Press* (pg. 79-114).

Moyer, G.R., Diaz-Ferguson, E., Hill, J.E. and Shea, C., 2014. Assessing environmental DNA detection in controlled lentic systems. *PLoS ONE*, 9(7), p.e103767.

Myers, N, Mittermeier, R.A., Mittermeier, C.G., Fonseca, G.A.B., Kent, J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.

Nathan, L. R. *et al.* 2014. Quantifying environmental DNA signals for aquatic invasive species across multiple detection platforms. *Environmental Science & Technology*., 48, pp. 12800-12806

Nathan, L.R. *et al.*, 2015. The use of environmental DNA in invasive species surveillance of the Great Lakes commercial bait trade. *Conservation Biology*, 29(2).

Nichols, R.V., Königsson, H., Danell, K. and Spong, G., 2012. Browsed twig environmental DNA: diagnostic PCR to identify ungulate species. Molecular Ecology Resources, 12(6), pp.983-989.

O'Donnell, J.L., Kelly, R.P., Lowell, N.C. and Port, J.A., 2016. Indexed PCR primers induce template-specific bias in large-scale DNA sequencing studies. *PLoS ONE*, 11(3), p.e0148698.

Ogram, A., Sayler, G.S. and Barkay, T., 1987. The extraction and purification of microbial DNA from sediments. *Journal of microbiological methods*, 7(2-3), pp.57-66.

Olds, B.P. *et al.*, 2016. Estimating species richness using environmental DNA. *Ecology and Evolution*, 6(12), pp.4214–4226.

Olson, Z.H., Briggler, J.T. and Williams, R.N., 2012. An eDNA approach to detect eastern hellbenders (*Cryptobranchus a. alleganiensis*) using samples of water. *Wildlife Research*, *39*(7), pp.629-636.

Ore, J., Elbaum, S., Burgin, A. *et al.* 2015. Autonomous aerial water sampling. In: Mejias L, Corke P, Roberts J (eds) *Field and service robotics*. Springer International Publishing, Switzerland, pp 137–151

Pace, N.R. 1997. A molecular view of microbial diversity and the biosphere. *Science*, 276 (1997), pp. 734-740

Palumbi, S., Hillis, D., Moritz, C., Mable, B. 1996. Nucleic acids II: the polymerase chain reaction, Molecular Systematics. 2nd Sunderland, MA Sinauer Assoc. pg. 205-247

Pawlowski, J. *et al.*, 2014. Environmental monitoring through protist next-generation sequencing metabarcoding: Assessing the impact of fish farming on benthic foraminifera communities. *Molecular Ecology Resources*, 14(6).

Pedersen, M.W., Overballe-Petersen, S., Ermini, L., Der Sarkissian, C., Haile, J., Hellstrom, M., Spens, J., Thomsen, P.F., Bohmann, K., Cappellini, E. and Schnell, I.B., 2015. Ancient and modern environmental DNA. *Philosophical Transactions, Royal Society B*, 370(1660), p.20130383.

Pegard, A., Miquel, C., Valentini, A., Coissac, E., Bouvier, F., François, D., Taberlet, P., Engel, E. and Pompanon, F., 2009. Universal DNA-based methods for assessing the diet of grazing livestock and wildlife from feces. *Journal of Agricultural and Food Chemistry*, *57*(13), pp.5700-5706.

Peh, K.S.H., 2010. Invasive species in Southeast Asia: the knowledge so far. *Biodiversity and Conservation*, 19(4), pp.1083-1099.

Pfleger, M.O. *et al.*, 2016. Saving the doomed: Using eDNA to aid in detection of rare sturgeon for conservation (Acipenseridae). *Global Ecology and Conservation*, 8.

Piaggio, A.J., Engeman, R.M., Hopken, M.W., Humphrey, J.S., Keacher, K.L., Bruce, W.E. and Avery, M.L., 2014. Detecting an elusive invasive species: a diagnostic PCR to detect Burmese python in Florida waters and an assessment of persistence of environmental DNA. *Molecular Ecology Resources*, *14*(2), pp.374-380.

Piggott, M.P., 2016. Evaluating the effects of laboratory protocols on eDNA detection probability for an endangered freshwater fish. *Ecology and Evolution*, 6(9).

Pilliod, D.S. *et al.*, 2013. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, 70(8).

Pilliod, D.S., Goldberg, C.S., Arkle, R.S. and Waits, L.P., 2014. Factors influencing detection of eDNA from a stream-dwelling amphibian. *Molecular Ecology Resources*, 14(1) pp.109-116.

Piñol, J., San Andrés, V., Clare, E.L., Mir, G. and Symondson, W.O.C., 2014. A pragmatic approach to the analysis of diets of generalist predators: the use of next-generation sequencing with no blocking probes. Molecular Ecology Resources, 14(1), pp.18-26.

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

Piñol, J., Senar, M.A. and Symondson, W.O., 2018. The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular ecology*.

Pochon, X., Bott, N.J., Smith, K.F. and Wood, S.A., 2013. Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. *PloS one*, 8(9), p.e73935.

Pochon, X., Zaiko, A., Fletcher, L.M., Laroche, O. and Wood, S.A., 2017. Wanted dead or alive? Using metabarcoding of environmental DNA and RNA to distinguish living assemblages for biosecurity applications. *PLoS ONE*, 12(11), p.e0187636.

Polz. M.F., Cavanaugh, C.M. 1998. Bias in Template-to-Product Ratios in Multitemplate PCR. Applied and Environmental Microbiology. 64(10):3724–3730.

Pompanon, F., Deagle, B.E., Symondson, W.O., Brown, D.S., Jarman, S.N. and Taberlet, P., 2012. Who is eating what: diet assessment using next generation sequencing. *Molecular ecology*, *21*(8), pp.1931-1950.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D. and Vogler, A.P., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic biology*, *55*(4), pp.595-609.

Pont, D., Rocle, M., Valentini, A., Civade, R., Jean, P., Maire, A., Roset, N., Schabuss, M., Zornig, H. and Dejean, T., 2018. Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Scientific reports*, 8(1), p.10361.

Porazinska, D.L., Giblin-Davis, R.M., Esquivel, A. *et al.* 2010. Ecometagenetics confirms high tropical rainforest nematode diversity. *Molecular Ecology*, 19, 5521–5530.

Porco, D., Rougerie, R., Deharveng, L. and Hebert, P., 2010. Coupling non-destructive DNA extraction and voucher retrieval for small soft-bodied Arthropods in a high-throughput context: the example of Collembola. *Molecular Ecology Resources*, *10*(6), pp.942-945.

Port, J.A. *et al.*, 2016. Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*, 25(2).

Porter, T.M. and Hajibabaei, M., 2018. Automated high throughput animal CO1 metabarcode classification. *Scientific reports*, 8(1), p.4226.

Poulsen, L.K., Ballard, G. and Stahl, D.A., 1993. Use of rRNA fluorescence in situ hybridization for measuring the activity of single cells in young and established biofilms. *Applied and environmental microbiology*, 59(5), pp.1354-1360.

Powers, T. 2004. Nematode molecular diagnostics: from bands to barcodes. *Annual Review of Phytopathology* 42: 367–383.

Preston, C.M., Harris, A., Ryan, J.P. *et al.* 2011. Underwater application of quantitative PCR on an ocean mooring. *PLoS ONE* 6:e22522.

Puillandre, N., Lambert, A., Brouillet, S. and Achaz, G., 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular ecology*, *21*(8), pp.1864-1877.

Qu, C. & Stewart, K.A., 2017. Comparing conservation monitoring approaches: traditional and environmental DNA tools for a critically endangered mammal.

Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., Mikhail, A. and Ouédraogo, N., 2016. Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589), p.228.

Raemy, M. and Ursenbacher, S., 2018. Detection of the European pond turtle (*Emys orbicularis*) by environmental DNA: is eDNA adequate for reptiles? *Brill*.

Ratnasingham, S. and Hebert, P.D., 2013. A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. *PloS one*, *8*(7), p.e66213.

Ratnasingham, S., and Herbert, P.D.N. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org).. *Molecular Ecology Resources*. 7(3), pp.355-364.

Razgour, O., Clare, E.L., Zeale, M.R., Hanmer, J., Schnell, I.B., Rasmussen, M., Gilbert, T.P. and Jones, G., 2011. High-throughput sequencing offers insight into mechanisms of resource partitioning in cryptic bat species. *Ecology and Evolution*, 1(4), pp.556-570.

Reed, J.Z., Tollit, D.J., Thompson, P.M. and Amos, W., 1997. Molecular scatology: the use of molecular genetic analysis to assign species, sex and individual identity to seal faeces. *Molecular ecology*, 6(3), pp.225-234.

Rees, H.C. *et al.*, 2014. The application of eDNA for monitoring of the Great Crested Newt in the UK. *Ecology and Evolution*, 4(21).

Rees, H.C. *et al.*, 2014. The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, 51(5).

Rees, H.C. *et al.*, 2015. Applications and limitations of measuring environmental DNA as indicators of the presence of aquatic animals. *Journal of Applied Ecology*, 52(4).

Reichhardt, T., 1999. It's sink or swim as a tidal wave of data approaches. *Nature*, 399(6736), pp.517-520.

Renshaw, M.A. *et al.*, 2015. The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol-chloroform-isoamyl alcohol DNA extraction. *Molecular Ecology Resources*, 15(1).

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P. and Coissac, E., 2011. ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, 39(21), pp.e145-e145.

Ribeiro, F.J., Przybylski, D., Yin, S., Sharpe, T., Gnerre, S., Abouelleil, A., Berlin, A.M., Montmayeur, A., Shea, T.P., Walker, B.J. and Young, S.K., 2012. Finished bacterial genomes from shotgun sequence data. *Genome research*, *22*(11), pp.2270-2277.

Robson, H.L., Noble, T.H., Saunders, R.J., Robson, S.K., Burrows, D.W. and 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(4), pp.922-932.

Rowley, J. *et al.* 2010. Impending conservation crisis for Southeast Asian amphibians. *Biology Letters* 6: 336–338.

Roy, M. *et al.*, 2017. Development of environmental DNA (eDNA) methods for detecting high-risk freshwater fishes in live trade in Canada. *Biological Invasions*, pp.1–16.

Salter, I. 2018. Seasonal variability in the persistence of dissolved environmental DNA (eDNA) in a marine system: The role of microbial nutrient limitation. *PLoS ONE*

Santos, S.R., Gutierrez-Rodriguez, C. and Coffroth, M.A., 2003. Phylogenetic identification of symbiotic dinoflagellates via length heteroplasmy in domain V of chloroplast large subunit (cp23S)—ribosomal DNA sequences. Marine Biotechnology, 5(2), pp.130-140.

Sarkissian, C., Ermini, L., Jónsson, H., Alekseev, A.N., Crubezy, E., Shapiro, B. and Orlando, L., 2014. Shotgun microbial profiling of fossil remains. *Molecular ecology*, *23*(7), pp.1780-1798.

Sassoubre, L.M. *et al.*, 2016. Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine Fish. *Environmental Science and Technology*, 50(19).

Sato, H., Sogo, Y., Doi, H. and Yamanaka, H., 2017. Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. Scientific reports, 7(1), p.14860.

Sato, Y., Miya, M., Fukunaga, T., Sado, T. and Iwasaki, W., 2018. MitoFish and MiFish pipeline: a mitochondrial genome database of fish with an analysis pipeline for environmental DNA metabarcoding. *Molecular biology and evolution.*

Saunders, D.L., Meeuwig, J.J. and Vincent, A.C.J., 2002. Freshwater protected areas: strategies for conservation. *Conservation Biology*, 16(1), pp.30-41.

Saxena, G. *et al.*, 2015. Ecogenomics reveals metals and land-use pressures on microbial communities in the waterways of a megacity. *Environmental Science and Technology*, 49(3), pp.1462–1471.

Schiebelhut, L.M. *et al.*, 2016. A comparison of DNA extraction methods for high-throughput DNA analyses. *Molecular Ecology Resources*.

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(4).

Schmidt, B.R., Kery, M., Ursenbacher, S., Hyman, O.J. and Collins, J.P., 2013. Site occupancy models in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. *Methods in Ecology and Evolution*, *4*(7), pp.646-653.

Schneider, G.F. and Dekker, C., 2012. DNA sequencing with nanopores. *Nature biotechnology*, *30*(4), pp.326-328.

Schneider, J. *et al.*, 2016. Detection of invasive mosquito vectors using environmental DNA (eDNA) from water samples. *PLoS ONE*, 11(9).

Schnell, I.B., Fraser, M., Willerslev, E. and Gilbert, M.T.P., 2010. Characterisation of insect and plant origins using DNA extracted from small volumes of bee honey. *Arthropod-Plant Interactions*, *4*(2), pp.107-116.

Schnell, I.B., Thomsen, P.F., Wilkinson, N., Rasmussen, M., Jensen, L.R., Willerslev, E., Bertelsen, M.F. and Gilbert, M.T.P., 2012. Screening mammal biodiversity using DNA from leeches. *Current biology*, *22*(8), pp.R262-R263.

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(6), pp.1289–1303.

Scott, D.A., 1989. A directory of Asian wetlands. IUCN, The World Conservation Union.

Seymour, M. *et al*. 2018. Acidity promotes degradation of multi-species eDNA in lotic mesocosms. *Communications Biology*. 1, 4.

Shaw, J.L.A. *et al.*, 2016. Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biological Conservation*, 197, pp.131–138.

Shogren, A.J. *et al.*, 2016. Modelling the transport of environmental DNA through a porous substrate using continuous flow-through column experiments. *Journal of The Royal Society Interface*, 13(119).

Shokralla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M. 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* 21:1794–1805.

Sigsgaard, E.E., Carl, H., Møller, P.R. and Thomsen, P.F., 2015. Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. *Biological Conservation*, 183, pp.46-52.

Sigsgaard, E.E. *et al.*, 2016. Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution*, 1(1), p.4.

Sigsgaard, E.E. *et al.*, 2017. Seawater environmental DNA reflects seasonality of a coastal fish community. *Marine Biology*, 164(6), p.128.

Simpfendorfer, C.A. *et al.*, 2016. Environmental DNA detects Critically Endangered largetooth sawfish in the wild. *Endangered Species Research*, 30(1).

Smart, A.S. *et al.*, 2015. Environmental DNA sampling is more sensitive than a traditional survey technique for detecting an aquatic invader. *Ecological Applications*, 25(7).

Smart, A.S. *et al.*, 2016. Assessing the cost-efficiency of environmental DNA sampling. *Methods in Ecology and Evolution*, 7(11).

Sneath, P.H. and Sokal, R.R., 1973. *Numerical taxonomy. The principles and practice of numerical classification.* San Francisco, W.H. Freeman and Company. ISBN 0716706970.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M. and Herndl, G.J., 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences*, 103(32), pp.12115-12120.

Soininen, E.M., Valentini, A., Coissac, E., Miquel, C., Gielly, L., Brochmann, C., Brysting, A.K., Sønstebø, J.H., Ims, R.A., Yoccoz, N.G. and Taberlet, P., 2009. Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. Frontiers in Zoology, 6(1), p.16.

Sokal, R.R., 1963. The principles and practice of numerical taxonomy. Taxon, pp.190-199.

Somervuo, P. *et al.*, 2017. Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding D. Warton, ed. *Methods in Ecology and Evolution*, 8(4), pp.398–407.

Son, M.S. and Taylor, R.K., 2011. Preparing DNA Libraries for Multiplexed Paired-End Deep Sequencing for Illumina GA Sequencers. *Current protocols in microbiology*, pp.1E-4.

Species Survival Commission, 2000. IUCN guidelines for the prevention of biodiversity loss caused by alien invasive species. Auckland, New Zealand: IUCN, Species Survival Commission, Invasive Species Specialist Group.

Spens, J. *et al.*, 2017. Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods in Ecology and Evolution*, 8(5), pp.635-645..

Stat, M., Huggett, M.J., Bernasconi, R., DiBattista, J.D., Berry, T.E., Newman, S.J., Harvey, E.S. and Bunce, M., 2017. Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, 7(1), p.12240.

Stewart, K. *et al.*, 2017. Using environmental DNA to assess population-wide spatiotemporal reserve use. *Conservation Biology*.

Stoeckle, B.C., Kuehn, R. & Geist, J., 2016. Environmental DNA as a monitoring tool for the endangered freshwater pearl mussel (Margaritifera margaritifera L.): a substitute for classical monitoring approaches? *Aquatic Conservation: Marine and Freshwater Ecosystems*, 26(6).

Stoeckle, M.Y. *et al.*, 2017. Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary H. Doi, ed. *PLoS ONE*, 12(4), p.e0175186.

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.

Sutherland, W.J. *et al.*, 2013. A horizon scan of global conservation issues for 2013. *Trends in Ecology & Evolution*, 28(1), pp.16–22.

Suyama Y, Kawamuro K, Kinoshita I, Yoshimura K, Tsumura Y, Takahara H. 1996. DNA sequence from a fossil pollen of Abies spp. from Pleistocene peat. *Genes & Genetic Systems* 71: 145–149.

Suzuki, M.T., Giovannoni, S.J. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology*. 62(2):625–630.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C. and Willerslev, E., 2006. Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. Nucleic acids research, 35(3), pp.e14-e14.

Taberlet, P., Coissac, E., Hajibabaei, M. and Rieseberg, L.H., 2012a. Environmental DNA. *Molecular ecology*, *21*(8), pp.1789-1793.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. and Willerslev, E., 2012b. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular ecology*, *21*(8), pp.2045-2050.

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H. and Kawabata, Z.I., 2012. Estimation of fish biomass using environmental DNA. *PLoS ONE*, *7*(4), p.e35868.

Takahara, T., Minamoto, T. and Doi, H., 2013. Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLoS ONE*, *8*(2), p.e56584.

Takahara, T., Minamoto, T. & Doi, H., 2015. Effects of sample processing on the detection rate of environmental DNA from the Common Carp (*Cyprinus carpio*). *Biological Conservation*, 183.

Taylor, P.G., 1996. Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. Molecular biology and evolution, 13(1), pp.283-285.

Thomas, A.C., Howard, J., Nguyen, P.L., Seimon, T.A. and Goldberg, C.S., 2018. ANDe™: A fully integrated environmental DNA sampling system. *Methods in Ecology and Evolution.*

Thomsen, P., Kielgast, J.O.S., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., Orlando, L. and Willerslev, E., 2012a. Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular ecology*, *21*(11), pp.2565-2573.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Møller, P.R., Rasmussen, M. and Willerslev, E., 2012b. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE*, *7*(8), p.e41732.

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

Thomsen, P.F. *et al.*, 2016. Environmental DNA from Seawater Samples Correlate with Trawl Catches of Subarctic, Deepwater Fishes A. R. Mahon, ed. *PLoS ONE*, 11(11), p.e0165252.

Torresdal, J.D., Farrell, A.D. & Goldberg, C.S., 2017. Environmental DNA detection of the golden tree frog (*Phytotriades auratus*) in bromeliads. *PLoS ONE*, 12(1).

Tréguier, A., Paillisson, J.M., Dejean, T., Valentini, A., Schlaepfer, M.A., Roussel, J.M., 2014. Environmental DNA surveillance for invertebrate species: advantages and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. Journal of Applied Ecology. 2014; 51:871–879. doi: 10.1111/ 1365-2664.12262.

Tsuji, S. *et al.*, 2017a. Water temperature-dependent degradation of environmental DNA and its relation to bacterial abundance. *PLoS ONE*, 12(4).

Tsuji, S., Yamanaka, H. & Minamoto, T., 2017b. Effects of water pH and proteinase K treatment on the yield of environmental DNA from water samples. *Limnology*, 18(1).

Tsuji, S., Iguchi, Y., Shibata, N., Teramura, I., Kitagawa, T. and Yamanaka, H., 2018. Real-time multiplex PCR for simultaneous detection of multiple species from environmental DNA: an application on two Japanese medaka species. *Scientific Reports*, 8(1), p.9138.

Tucker, A.J. *et al.*, 2016. A sensitive environmental DNA (eDNA) assay leads to new insights on Ruffe (*Gymnocephalus cernua*) spread in North America. *Biological Invasions*, 18(11).

Turner, C.R. *et al.*, 2014a. Improved methods for capture, extraction, and quantitative assay of environmental DNA from Asian bigheaded carp (hypophthalmichthys spp.). *PLoS ONE*, 9(12).

Turner, C.R. *et al.*, 2014b. Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution*, 5(7).

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.

Uchii, K., Yamanaka, H. and Minamoto, T., 2017. Distinct seasonal migration patterns of Japanese native and non-native genotypes of common carp estimated by environmental DNA. *Ecology and Evolution*.

Ushio, M. *et al.*, 2017. Environmental DNA enables detection of terrestrial mammals from forest pond water. *Molecular Ecology Resources*.

Valentini, A., Pompanon, F. and Taberlet, P., 2009. DNA barcoding for ecologists. *Trends in Ecology & Evolution*, *24*(2), pp.110-117.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F. and Gaboriaud, C., 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, 25(4), pp.929-942.

Van Oosterzee P. 1997. Where worlds collide, The Wallace Line. Ithaca (NY) *Cornell University Press*.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W. and Fouts, D.E., 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 304(5667), pp.66-74.

Vervoort, M.T.W., Vonk, J.A., Mooijman, P.J.W. *et al.* 2012. SSU ribosomal DNA-based monitoring of nematode assemblages reveals distinct seasonal fluctuations within evolutionary heterogeneous feeding guilds. *PLoS ONE* 7:e47555.

Vestheim, H. and Jarman, S.N., 2008. Blocking primers to enhance PCR amplification of rare sequences in mixed samples–a case study on prey DNA in Antarctic krill stomachs. *Frontiers in zoology*, *5*(1), p.12.

Vörös, J., Márton, O., Schmidt, B.R., Gál, J.T. and Jelić, D., 2017. Surveying Europe's only cave-dwelling chordate species (*Proteus anguinus*) using environmental DNA. *PLoS ONE*, 12(1), p.e0170945.

Vörösmarty, C.J., *et al.* 2010. Global threats to human water security and river biodiversity. *Nature* 467: 555–561.

Walker, S.F., Salas, M.B., Jenkins, D., Garner, T.W., Cunningham, A.A., Hyatt, A.D., Bosch, J. and Fisher, M.C., 2007. Environmental detection of Batrachochytrium dendrobatidis in a temperate climate. *Diseases of aquatic organisms*, *77*(2), pp.105-112.

Wallace, A.R., 1869. *The Malay Archipelago: the land of the orang-utan and the bird of paradise; a narrative of travel, with studies of man and nature*. Courier Corporation.

Wang, J., Li, G. and Wang, J., 1981. *The Early Tertiary fossil fishes from Sanshui and its adjacent basin, Guangdong*. *Palaeontologia Sinica*, No. 160, Serie C(22) (nieuwe)

Weltz, K. *et al*. 2017. Application of environmental DNA to detect an endangered marine skate species in the wild. PLoS One 12, e0178124.

Ward, R.D., Zemlak, T.S., Innes, B.H. *et al.* 2005. A start to DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London*. Series B, Biological Sciences, 360, 1847–1857.

Wegleitner, B.J. *et al.*, 2015. Long duration, room temperature preservation of filtered eDNA samples. *Conservation Genetics Resources*, 7(4).

Wiemers, M. and Fiedler, K., 2007. Does the DNA barcoding gap exist?–a case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in zoology*, 4(1), p.8.

Wilcox, T.M., McKelvey, K.S., Young, M.K., Jane, S.F., Lowe, W.H., Whiteley, A.R. and Schwartz, M.K., 2013. Robust detection of rare species using environmental DNA: the importance of primer specificity. *PLoS ONE*, *8*(3), p.e59520.

Wilcox, T.M., Schwartz, M.K., McKelvey, K.S. *et al.* 2014. A blocking primer increases specificity in environmental DNA detection of bull trout (*Salvelinus confluentus*). *Conservation Genetics Resources* 6:283–284.

Wilcox, T.M. *et al.*, 2015. Environmental DNA particle size distribution from Brook Trout (*Salvelinus fontinalis*). *Conservation Genetics Resources*, 7(3)

Wilcox, T.M., Zarn, K.E., Piggott, M.P., Young, M.K., McKelvey, K.S. and Schwartz, M.K. 2018. Capture enrichment of aquatic environmental DNA: A first proof of concept. *Molecular Ecology Resources.*

Wilkinson, S.P., Davy, S.K., Bunce, M. and Stat, M., 2018. Taxonomic identification of environmental DNA with informatic sequence classification trees (No. e26812v1). PeerJ Preprints.

Willerslev, E., Hansen, A.J., Binladen, J., Brand, T.B., Gilbert, M.T.P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D.A. and Cooper, A., 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, *300*(5620), pp.791-795.

Willerslev, E., Hansen, A.J. and Poinar, H.N., 2004. Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology & Evolution*, *19*(3), pp.141-147.

Willerslev, E., Cappellini, E., Boomsma, W., Nielsen, R., Hebsgaard, M.B., Brand, T.B., Hofreiter, M., Bunce, M., Poinar, H.N., Dahl-Jensen, D. and Johnsen, S., 2007. Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science*, *317*(5834), pp.111-114.

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M.E., Lorenzen, E.D., Vestergård, M., Gussarova, G., Haile, J. and Craine, J., 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, *506* (7486), p.47.

Williams, K.E., Huyvaert, K.P. & Piaggio, A.J., 2016. No filters, no fridges: a method for preservation of water samples for eDNA analysis. *BMC Research Notes*, 9(1)

Williams, K.E. *et al.*, 2017. Clearing muddied waters: Capture of environmental DNA from turbid waters H. Doi, ed. *PLoS ONE*, 12(7), p.e0179282Williams, K.E., Huyvaert, K.P., Vercauteren, K.C., Davis, A.J. and Piaggio, A.J., 2018. Detection and persistence of environmental DNA from an invasive, terrestrial mammal. *Ecology and evolution.*

Woese, C.R. 1987.Bacterial evolution Microbiology. Rev., 51, p. 221

Woodruff D.S. 2010. Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. *Biodiversity Conservation*. Vol. 19 (pg. 919-941)

Wu, Q., Kawano, K., Uehara, Y., Okuda, N., Hongo, M., Tsuji, S., Yamanaka, H. and Minamoto, T., 2018. Environmental DNA reveals nonmigratory individuals of Palaemon paucidens overwintering in Lake Biwa shallow waters. *Freshwater Science*, 37(2).

Xie, Y. *et al.*, 2016. Environmental DNA metabarcoding reveals primary chemical contaminants in freshwater sediments from different land-use types. *Chemosphere*, 172, pp.201–209.

Yamamoto, S. *et al.*, 2017. Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific Reports*, 7.

Yamanaka, H. *et al.*, 2016a. On-site filtration of water samples for environmental DNA analysis to avoid DNA degradation during transportation. *Ecological Research*, 31(6).

Yamanaka, H. and Minamoto, T., 2016b. The use of environmental DNA of fishes as an efficient method of determining habitat connectivity. *Ecological Indicators*, 62, pp.147-153.

Yamanaka, H. *et al.*, 2017. A simple method for preserving environmental DNA in water samples at ambient temperature by addition of cationic surfactant. *Limnology*, 18(2), pp.233–241.

Yoccoz, N.G., 2012a. The future of environmental DNA in ecology. *Molecular Ecology*, 21(8), pp.2031–2038.

Yoccoz, N.G., Bråthen, K.A., Gielly, L., Haile, J., Edwards, M.E., Goslar, T., Von Stedingk, H., Brysting, A.K., Coissac, E., Pompanon, F. and Sønstebø, J.H., 2012b. DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology*, 21(15), pp.3647-3655.

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.

Yule, C. M. "Freshwater environments." Freshwater Invertebrates of the Malaysian Region. Academy of Sciences Malaysia Publishers, Malaysia (2004).

Zepeda Mendoza, M.L., Sicheritz-Pontén, T. and Gilbert, M.T.P., 2015. Environmental genes and genomes: understanding the differences and challenges in the approaches and software for their analyses. *Briefings in bioinformatics*, 16(5), pp.745-758.

Zhan, A., Xiong, W., He, S. and MacIsaac, H.J., 2014. Influence of artifact removal on rare species recovery in natural complex communities using high-throughput sequencing. *PloS one*, 9(5), p.e96928.

Zhang, J., Kapli, P., Pavlidis, P. and Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, *29*(22), pp.2869-2876.

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., Tang, M., Fu, R., Li, J. and Huang, Q., 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *Gigascience*, *2*(1), p.4.

Zinger, L., Gobet, A. and Pommier, T., 2012. Two decades of describing the unseen majority of aquatic microbial diversity. *Molecular Ecology*, *21*(8), pp.1878-1896.

1.19 Online References

ADAS. 2017. *EDNA ANALYSIS FOR GREAT CRESTED NEWT*. Service. ADAS. http://www.adas.uk/Service/edna-analysis-for-great-crested-newt

American Meteorological Society. 2012. *Meteorology Glossary*. Freshwater. http://glossary.ametsoc.org/wiki/Freshwater

Amos, J. 2012. Mapping Earth's surface in 3D. Science & Environment. BBC News. http://www.bbc.co.uk/news/science-environment-16578176

BOLDSYSTEMS. 2017a. TAXONOMY. Kingdoms of Life Being Barcoded. http://www.boldsystems.org/index.php/TaxBrowser_Home Accessed on 27/09/2017

BOLDSYSTEMS. 2017b. Barcodes per site. http://v3.boldsystems.org/

CALeDNA. 2018. Our Mission. About the Project. Accessed at: http://www.ucedna.com/implications-for-conservation/

Atlas of Living Australia. 2017. *The Atlas of Living Australia is a collaborative, national project that aggregates biodiversity data from multiple sources and makes it freely available and usable online*. https://www.ala.org.au/

Eisen, J. 2012. *Referring to 16S surveys as "metagenomics" is misleading and annoying #badomics #OmicMimicry*. The Tree of Life. https://phylogenomics.blogspot.co.uk/2012/08/referring-to-16s-surveys-as.html

Eschmeyer, W. N., Fricke, R. and van der Laan R. (eds). *SPECIES BY FAMILY/SUBFAMILY IN THE*. CATALOG OF FISHES: GENERA, SPECIES, REFERENCES. http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp Electronic version accessed 01/09/2017

Freshwater Habitats Trust. 2017. *PondNet Results 2015, 2016 & 2017*. Projects and Surveys. Pond Net. Freshwater Habitats Trust. https://freshwaterhabitats.org.uk/projects/pondnet/pondnet-results-2015-2016-2017/

GOV.UK. 2017. *Great crested newts: surveys and mitigation for development projects*. Guidance. GOV.UK. https://www.gov.uk/guidance/great-crested-newts-surveys-and-mitigation-for-development-projects

GBIF. 2018. Free and open access to biodiversity data. Global Biodiversity Information Facility. Accessed at: https://www.gbif.org/

IUCN. 2017. *The IUCN Red List of Threatened Species*. http://www.iucnredlist.org/

Mega Fishing Thailand, 2017. *INTRODUCED*. http://megafishingthailand.com/fish-species-in-thailand/introduced/

Schloss, P. D. 2018. *Pre-print Review of 97% Identity Threshold Manuscript*. The Academic Hermit. http://www.academichermit.com/2017/10/11/Review-of-Updating-the-97-identity-threshold.html

Soala Working Group. 2013. *Conservation Through Collaboration: Proceedings of the 3rd Meeting of the Saola Working Group 2013*. Vientiane, Law PDR, 3-7 June 2013. http://www.cepf.net/SiteCollectionDocuments/indo_burma/TechnicalReport_54444_GWC_SWG_ThirdMeeting_Proceedings.pdf

Spyboat. 2017. *Aquatic drone for inspection and environmental sampling*. AQUATIC DRONE. SPYBOAT. http://www.spyboat-technologies.com/

The World Conservation Union. (2010). IUCN Red List of Threatened Species. Summary Statistics for Globally Threatened Species. Table 1: Numbers of threatened species by major groups of organisms (1996–2010).

United Nations Statistics Division. 2012. http://unstats.un.org/unsd/methods/m49/m49regin.htm

Watson, M. 2014. *You're probably not doing metagenomics*. Opinionomics. http://biomickwatson.wordpress.com/2014/01/12/youre-probably-not-doing-metagenomics//

Woldometers. 2018. South-Eastern Asia Population (LIVE). South-Eastern Asia. Asia. World. Population. Accessed at: http://www.worldometers.info/world-population/south-eastern-asia-population/

WWF (2012). (http://wwf.panda.org/about_our_earth/all_publications/living_planet_report/living_planet_report_graphics/lpi_interactive/)

Chapter 2

Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter

2.1 Abstract

1. Aqueous environmental DNA (eDNA) is an emerging efficient non-invasive tool for species inventory studies. To maximize performance of downstream quantitative PCR (qPCR) and next-generation sequencing (NGS) applications, quality and quantity of the starting material is crucial, calling for optimized capture, storage and extraction techniques of eDNA. Previous comparative studies for eDNA capture/storage have tested precipitation and 'open' filters. However, practical 'enclosed' filters which reduce unnecessary handling have not been included. Here, we fill this gap by comparing a filter capsule (Sterivex-GP polyethersulfone, pore size 0.22 µm, hereafter called SX) with commonly used methods.
2. Our experimental set-up, covering altogether 41 treatments combining capture by precipitation or filtration with different preservation techniques and storage times, sampled one single lake (and a fish-free control pond). We selected documented capture methods that have successfully targeted a wide range of fauna. The eDNA was extracted using an optimized protocol modified from the DNeasy® Blood & Tissue kit (Qiagen). We measured total eDNA concentrations and Cq-values (cycles used for DNA quantification by qPCR) to target specific mtDNA cytochrome *b* (cyt *b*) sequences in two local keystone fish species.
3. SX yielded higher amounts of total eDNA along with lower Cq-values than polycarbonate track-etched filters(PCTE), glass fibre filters (GF) or ethanol precipitation (EP). SX also generated lower Cq-values than cellulose nitrate filters (CN) for one of the target species. DNA integrity of SX samples did not decrease significantly after 2 weeks of storage in contrast to GF and PCTE. Adding preservative before storage improved SX results.
4. In conclusion, we recommend SX filters (originally designed for filtering micro-organisms) as an efficient capture method for sampling macrobial eDNA. Ethanol or Longmire's buffer preservation of SX immediately after filtration is recommended. Preserved SX capsules may be stored at room temperature for at least 2 weeks without significant degradation. Reduced handling and less exposure to outside stress compared with other filters may contribute to better eDNA results. SX capsules are easily transported and enable eDNA sampling in remote and harsh field conditions as samples can be filtered/preserved on site.

Key-words: capsule, eDNA capture, environmental DNA, extraction, filter, monitoring, quantitative PCR, species-specific detection, water sampling method

2.2 Introduction

The realization that DNA from macrobiota can be obtained from environmental samples (environmental DNA, eDNA) started with excrements (Höss *et al*. 1992) and sediments (Willerslev *et al*. 2003). Over the last decade, the potential of aqueous eDNA to identify a wide range of plants and animals from a small volume of water has been realized (Martellini, Payment & Villemur 2005; Thomsen *et al*. 2012; Rees *et al*. 2014). Aqueous eDNA is an emerging increasingly sensitive technique for revealing species distributions (e.g. Jane *et al*. 2015; Valentini *et al*. 2016), early detection of invasive species (e.g. Smart *et al*. 2015; Simmons *et al*. 2016) and monitoring rare and/or threatened species for conservation (e.g. Zhan *et al*. 2013; McKee *et al*. 2015). Aqueous eDNA monitoring provides possibilities to upscale species distribution surveys considerably, because much less effort in time and resources are required compared to conventional methods (Dejean *et al*. 2012; Davy, Kidd & Wilson 2015). Based on literature searches, we catalogue 49 studies successfully applying eDNA from water samples to detect macro-organisms in aquatic ecosystems, published between January 2005 and March 2015 (when this study was initiated; Table S1, Supporting Information). To our knowledge, 39 additional empirical studies were published since then, indicating a rapid rise of interest in this research area (Table S2). The field of eDNA is still evolving, and a consensus of capture, storage and extraction methods has not yet been reached (Goldberg, Strickler & Pilliod 2015; Tables S1 and S2). In fact, the diversity of methods is almost as high as the number of research groups investigating this fairly new field of research. To ensure reliable results of downstream applications such as quantitative PCR (qPCR) and next-generation sequencing (NGS), the quantity and quality of the starting material is crucial. From our eDNA laboratory experience, we find that a modified easy-to-follow extraction protocol resulting in high yields is needed. Based on eDNA studies published so far (Tables S1 and S2), we identify three pre-PCR key issues that hold opportunities for improvement: (i) capturing sufficient quantities of eDNA as quite a few studies report low amounts of captured total eDNA, (ii) effectively preserving eDNA samples before extraction and (iii) lowering contamination risks from collection to extraction of eDNA. Comparative studies on aqueous eDNA capture and storage techniques (i.e. optimal ways of preserving the eDNA captured on the filters until extraction; e.g. Renshaw *et al*. 2015) were based on the so-called 'open filters' (requiring handling, a filter funnel and a vacuum pump; e.g. Liang & Keeley 2013; Turner *et al*. 2014b) and ethanol precipitation (EP; e.g. Piaggio *et al*. 2014; Deiner *et al*. 2015). However, no enclosed filters were included in

previous comparative assays. The Sterivex-GP capsule filter (SX), with a polyethersulfone membrane, is a standard method for characterizing microbial communities (Chestnut *et al*. 2014) and for removing pathogens from water as the organisms are captured on the filter membranes. To our knowledge, only two published aqueous eDNA studies have used this filter to detect aquatic macroorganisms (fish detection: Keskin 2014; Bergman *et al*. 2016), and the technique has been successful to detect a wide range of aquatic macro-organisms in Denmark and Belgium (M. Hellström, M.E. Sengupta, S.W. Knudsen, D. Halfmarten. unpublished, S1). The SX filter is enclosed in a capsule, which reduces handling. A water sample can easily be filtered in the field, saving time and facilitating fixation of the eDNA immediately after capture. Additionally, downstream DNA extraction takes place within the filter capsules with no need for the membrane to be removed or handled. We therefore test the performance of SX compared to other more frequently used eDNA capture methods (Table S1), under different storage conditions, in an effort to address issues 1–3 above. To date, there are no studies comparing SX to other capture methods and multiple storage treatments. We aim to fill this gap, with an experimental study comparing SX with four other capture methods in a set-up with five typical storage treatments and three different storage times (up to 2 weeks). The tested open filter materials polycarbonate, cellulose nitrate and glass fibre (GF) and the range of tested pore sizes (0.2–0.6 µm) are typical of previous studies (Tables S1 and S2). We used an optimized extraction protocol based on a commercial kit to increase eDNA yields.

2.2.1 Hypotheses

To evaluate the usefulness of the SX and preservation buffers in comparison with typically used methods (Tables S1 and S2), we test the following H_0 hypotheses:

$H_0$1. *Capture method*: SX is equally effective as other tested eDNA capturing techniques in regard to DNA quantity and quality measured as the total extracted eDNA concentration [$eDNA_{tot}$] and as Cq-values (quantification cycles, sensu Bustin *et al*. 2009) from two species-specific qPCR assays.

$H_0$2a. *Storage preservative*: Storing filters with a preservation buffer does not affect qPCR amplification compared to immediate extraction or freezing at -20 °C (no buffer added).

$H_0$2b. *Storage time*: There is no significant difference in eDNA quality over time between SX and the other tested capturing techniques.

$H_0$3. *Contamination*: There is no significant difference between SX and the other tested capture techniques in occurrence of false positives.

To test these hypotheses, we use an experimental set-up with subsampling a single large homogenous sample of water from a Danish lake. Subsamples are subjected to different eDNA capture methods within the same day followed by different storage treatments. A control site (fish-free pond) is sampled using the same set-up. Each capture and storage treatment is assessed using concentration of total eDNA as well as species specific qPCR assays targeting pike *Esox lucius L.* and perch
Perca fluviatilis L.

By testing H_0 hypotheses (1–3), the multiple opportunities for optimization of eDNA surveys held by the use of SX may be empirically evaluated. Based on the results, we suggest recommendations for improved capture, storage and extraction to use for aqueous eDNA, taking remote and harsh field conditions into consideration.

2.3 Materials and methods

2.3.1 Study sites

We chose Gentofte Lake, Denmark (N55.7435°, E12.5348°), as the study site and a fish-free pond in Copenhagen botanical garden as a negative field control (N55.6875°, E12.5746°). Gentofte Lake (26 ha) is an alkaline clear water (Appendix S2) harbouring a wide range of fish species, including pike and perch.

2.3.2 Water collection

We retrieved 130 L of water from Gentofte Lake on 17 March 2015. The water (4 °C) was collected at c. 30 points along c. 100 m of shoreline close to the outlet of the lake. Additionally, we collected 40 L of water from the control pond on 21 March 2015. The water was collected in sterilized 5-L buckets which prior to sampling were soaked in bleach (5%) for 10 min, and then rinsed with laboratory-grade ethanol (70%). The containers were soaked repeatedly in lake water at a location away from the collection point. Nitrile gloves were used during cleaning, collection and filtration.

2.3.3 Capture and storage

We carried out 41 different treatment combinations of the water sample in total (Table 1, Fig. S1). We used five capture techniques, five storage methods and three time regimes. All treatments were performed in triplicate. Apart from an in-house modified SX procedure (see Fig. 1), the capture and storage methods were based on published sources (Table S1). The capture methods (hereafter referred to with their abbreviations in square brackets) were as follows: (i) ethanol precipitation [EP] (Ficetola *et al*. 2008), (ii) mixed cellulose esters membrane filters including cellulose nitrate and cellulose acetate [CN]; Advantec 47 mm diameter 0.45 lm pore size (Toyo Roshi Kaisha, Ltd., Tokyo, Japan), (iii) polycarbonate track-etched filters [PCTE]; Whatman Nucleopore Membrane 47 mm diameter 0.2 lm pore size (Merck KGaA, Darmstadt, Germany)], (iv) glass fibre [GF] membrane filters; Advantec GA-55 47 mm diameter 0.6 lm pore size (Toyo Roshi Kaisha, Ltd., Tokyo, Japan) and (v) sterivex-GP capsule filters [SX]; polyethersulfone 0.22 µm pore size with luer-lock outlet (Merck KGaA)]. Further downstream, SX was divided into an extraction from the filter within the capsule ($SX_{CAPSULE}$), after removal of the storage buffer, and an extraction from

the removed preservation buffer within a centrifuge tube (SX_{TUBE}; see DNA extraction section below). The different storage methods were as follows: (i) ethanol 99% 200 proof at room temperature (RT), Molecular Biology Grade (Thermo Fisher Scientific Inc. Waltham, MA, USA); (ii) Longmire's buffer at RT (Longmire's; Longmire, Maltbie & Baker 1997); (iii) RNAlater at RT (RNA Stabilization Reagent; QIAGEN, Stockach, Germany); (iv) no buffer, frozen at 20 °C; and (v) no buffer, refrigerated at 8–10 °C. The three time regimes between filtration and extractions were (i) within 5 hours (5 h), (ii) within 24 h and (iii) after 2 weeks. Each treatment (n = 41) was performed in triplicate. For each filter replicate, 1 L of lake water was processed (0.015 L for EP). For each capture–storage treatment, we included one negative control without lake water. Additionally, 1 L tap water was run through each filter (0.015 L for EP) as a control to detect potential contamination from the filtration facilities. For the control pond, one sample per capture–storage treatment was processed (n = 23). We captured eDNA from 155 subsamples and negative controls altogether. The water samples were filtered or ethanol-precipitated by a team of 10 researchers and the replicates of each treatment started at different times to avoid temporal bias of filtrations. Prior to DNA capture, bench surfaces and all equipment were wiped with bleach (5%) and laboratory-grade ethanol (70%). Prior to each collection of subsamples, the water was mixed thoroughly in the 130-L container. For the open membrane filter (GF, CN and PCTE), 1 L water samples were vacuum-filtered (c. 15–30 min) using Nalgene 250-mL sterile disposable test filter funnels (Thermo Fisher Scientific Inc. USA). The filters were removed from the funnel with forceps and then placed in 5- mL DNA LoBind® centrifuge tubes (Eppendorf AG, Hamburg, Germany) that were either empty (if the time regime was 5 h or the storage method was freezing) or contained preservation buffer. For all treatments and downstream applications, Eppendorf DNA LoBind® tubes were used in order to avoid up to 50% retention of DNA by the plastic, which is a documented problem especially for short DNA fragments (Gaillard & Strauss 1998; Ellison *et al*. 2006). For the SX filters, 1 L of water was slowly (c. 10 min to avoid tearing of filters, following manufacturer's recommendations) pushed through each filter capsule using a prepacked sterile 50-mL luer-lock syringe. Remaining water in the SX was removed by pushing air through the filter until dry, also using the syringe. The outlet ends of the filters were closed with MoBio outlet caps (MOBIO Laboratories, QIAGEN) and 2 mL preservation buffer was pipetted to the inlet end using filter tips. The inlet ends were closed with inlet caps (MOBIO Laboratories, QIAGEN) and both ends were sealed with parafilm where after the capsules were inverted vigorously. The frozen samples and the (5 h) and (24 h) EP samples were placed at -20 °C until

extraction, while the non-treated samples (5 h) were placed in a refrigerator and extracted directly after the filtering session. Samples containing buffers were stored at RT until processed. The (2 weeks) EP samples were frozen for 24 h prior to extraction to allow for precipitation. In total, we processed 96.135 L of water from the lake (32 treatments 9 3 replicates 9 1 L + 3 EP treatments 9 3 replicates 9 0.015 L) and 20.045 L of water from the control pond (20 treatments 9 1 replicate 9 1 L +3 EP treatments9 1 replicates 9 0.015 L; Table 1).

2.3.4 Molecular laboratory conditions

DNA extractions and qPCR assays took place in the laboratories at the Centre for GeoGenetics, University of Copenhagen, Denmark. The facilities are designed for handling environmental samples requiring the most stringent precautions to avoid contamination. Pre-PCR, extraction and PCR facilities are located in separate designated rooms with positive air pressure. Laboratory coats are changed between rooms. Prior to any work in the laboratory, all surfaces are washed with 5% bleach and 70% ethanol. After completing extractions involving guanidiumthiocyanate, surfaces are washed with 70% ethanol (to avoid reactions between chlorine in the bleach and guanidiumthiocyanate in two of the buffers provided with the Qiagen kit), 5% bleach and then 70% ethanol. All extractions of eDNA took place in laminar flow hoods which were UV-treated before and after extractions. Every night, the entire facilities are automatically UV treated for a 2-h period.

2.3.5 DNA extraction

We extracted the eDNA using the extraction protocol outlined in Fig. 1 and Appendix S1. The SX filters containing preservation buffers underwent two extractions, one extraction from the buffer and one extraction within the filter capsule after it had been emptied of buffer (hereafter referred to as SX_{TUBE} and $SX_{CAPSULE}$). Altogether, 179 (24 SX_{TUBE} + 155 (see 'Capture and storage' section above) samples from the study lake and the control pond were extracted. We measured [$eDNA_{tot}$] in each extraction using a Qubit 1.0 fluorometer (Thermo Fisher Scientific Inc.) applying the high-sensitivity assay for dsDNA (Life Technologies, Carlsbad, CA, USA).

2.3.6 Quantitative PCR

For the qPCR assays (e.g. Wilcox *et al*. 2013), two species-specific Taq- Man primers/probe sets were used targeting 84 and 89 base pair fragments of the mitochondrial cytochrome *b*

(cyt *b*) gene in pike and perch, respectively (Table S3). Species specificity of the assays was tested on extracted DNA from non-target species (Table S3) using the qPCR set-up described below. These non-target species did not generate any amplification signals. The optimal ratio of probe: primer concentration was tested prior to the study. The final PCR set-up to detect the target species was as follows: pike – 5 µL template DNA, 12.5 µL TaqMan Environmental Master Mix 2.0 (Life Technologies), 3 µL forward primer (10 µM), 2 µL reverse primer (10 µM) and 3 µL probe (2.5 µM); and perch – 5 µL template DNA, 12.5 µL TaqMan Environmental Master Mix 2.0 (Life Technologies), 0.5 µL forward primer (10 µM), 2.5 µL reverse primer (10 µM), 3 µL probe (2.5 µM) and 1.5 µL UV-treated laboratory-grade water. The TaqMan qPCRs were performed on a Stratagene Mx3005P (Thermo Fisher Scientific Inc.) using thermal cycling parameters of 50 °C (5 min), 95 °C (10 min) followed by 50 cycles of 95 °C (30 s) and 60 °C (1 min). For each plate, no-template controls (NTCs) and positive/negative tissue extracts were run alongside the samples. All filtering and extraction negatives were included in the qPCR assays. Additional qPCR replicates were run in order to detect effects of freezing and thawing of the samples. To check for PCR inhibition in the lake, separate qPCR assays for both species following the protocols above were performed in a dilution series (1:1, 1:2, 1:10 and 1:20) of extracted DNA on four samples replicated twice plus two positive and two negative controls to determine any deviation of the amplification curves. The dilution series did not indicate inhibition.

2.3.7 Data analysis

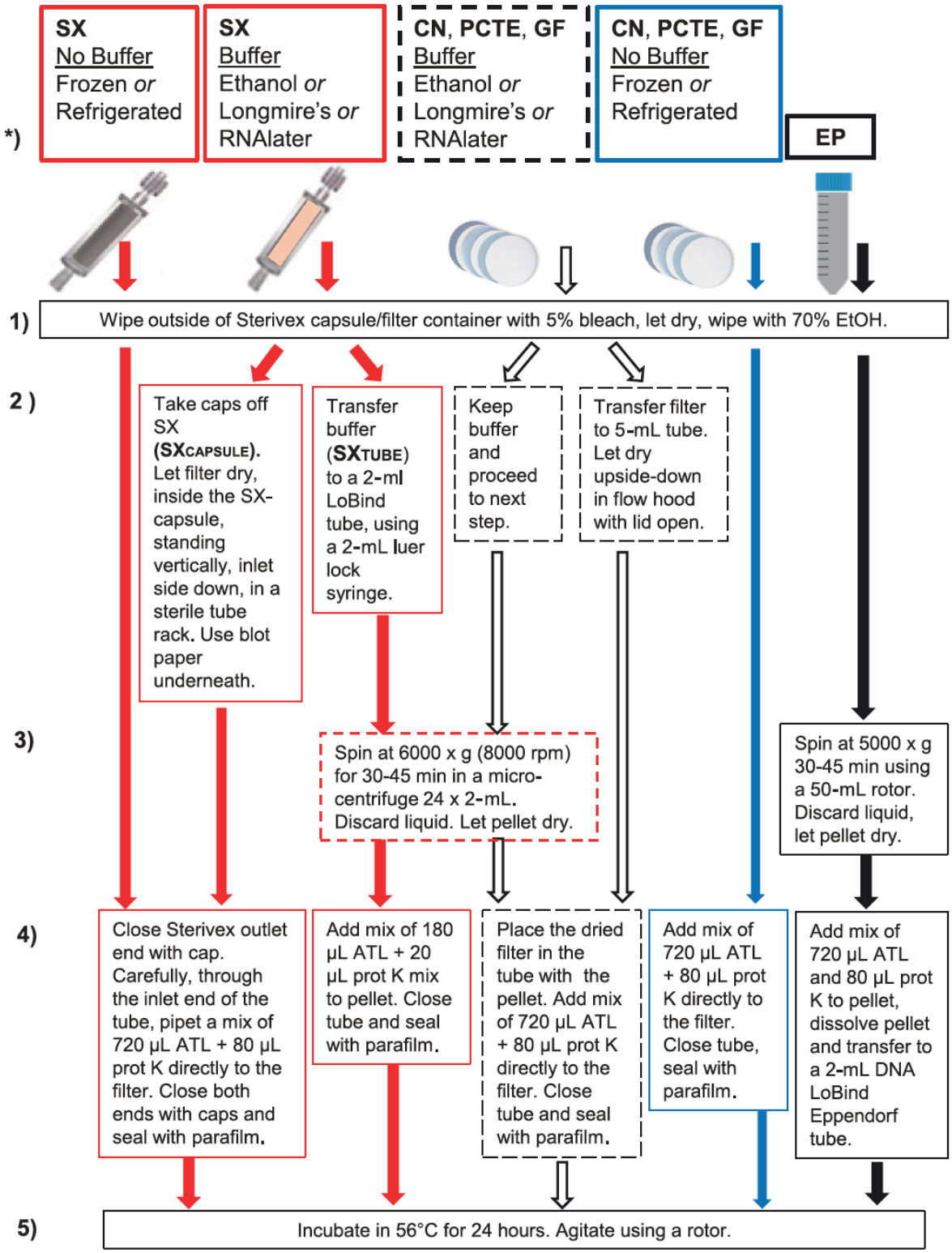
To compare detection probability (i.e. diagnostic sensitivity) between eDNA capture methods, the proportion of positive qPCR replicates was calculated for each target species. Positive samples were analysed using multivariate decision trees and univariate tests of 'no-effect' null hypotheses. To explore the effect of capture and storage on qPCR Cq values, Chi-square Automatic Interaction Detector (CHAID) decision tree was used. CHAID is a nonparametric tree-building method that can handle multivariate categorically induced quantitative responses (IBM Corp. (2013)). It defines optimal multiway splits and adjusts for Bonferroni. The main advantage of this approach is to analyse a data set all-in-one (rather than manually splitting the data into user-selected subgroups and thereafter choosing and performing multiple tests). The approach offers a number of other advantages including its ability to handle categorical (ordered, nominal) data types well and to model nonlinear

relationships without having to specify a priori the form of the interactions. A CHAID tree produces an overview, grouping or singling out the factors that predict the variation in theresponse variable. Categorical variables (capture method, storage treatment and storage time) were used as model predictors, and Cq-value from qPCR was set as the response target. Two trees were generated: the first targeting perch and the second pike. Tree depth, that is the maximum number of branching levels, was set to two (realized from ten 50/50 split validations) to reduce overfitting. For a univariate test of H_0 (1–2a, b), first a Wilcoxon signed-rank test for paired samples was applied to determine whether [$eDNA_{tot}$] and Cq values attained using $SX_{CAPSULE}$ differ significantly, from any of the other tested capture methods (CN, GF, PCTE, EP and SX_{TUBE}). Secondly, SX, GF and PCTE filter results were tested for signs of eDNA degradation over time, that is detecting any significant difference in Cq-values or [$eDNA_{tot}$] between 24 h and 2 weeks of storage. Wilcoxon signed-rank test was used as data exhibited non-normal distributions. Thirdly, guided by results from the CHAID trees, results from $SX_{CAPSULE}$ stored in ethanol or Longmire's were tested (Mann–Whitney) for differences in Cq-value against $SX_{CAPSULE}$ without preservation buffer. The CN filter group was reduced, as the planned 1-day storage treatment was omitted due to filtering time constraints. The mean difference in Cq-value and associated 95% CI of all qPCR replicates was calculated. All statistical analyses were performed using SPSS IBMCorp. (2013).

Table 2.1. Outline of the number of samples processed per capture and storage treatment. Negative control pond in parentheses. Sterivex, eDNA extraction within capsule ($SX_{CAPSULE}$); Sterivex, eDNA extraction from buffer in tube outside capsule (SX_{TUBE}).

		Storage								
		Refrigerated	*Frozen*	*Ethanol*	*Longmire's*	*RNAlater*	*Frozen*	*Ethanol*	*Longmire's*	*RNAlater*
Capture	**Sum**		*24 h*				*2 weeks*			
$SX_{CAPSULE}$	27 (5)	3 (1)	3 (1)	3 (1)	3 (1)	3 (1)	3	3	3	3
SX_{TUBE}	18 (3)			3 (1)	3 (1)	3 (1)		3	3	3
Cellulose nitrate	15 (5)	3 (1)	(1)	(1)	(1)	(1)	3	3	3	3
Glass fibre	27 (5)	3 (1)	3 (1)	3 (1)	3 (1)	3 (1)	3	3	3	3
Polycarbonate	27 (5)	3 (1)	3 (1)	3 (1)	3 (1)	3 (1)	3	3	3	3
Precipitation	9 (3)	3		3 (1)				3		
Total	123 (26)									

DNA extraction: DAY 1



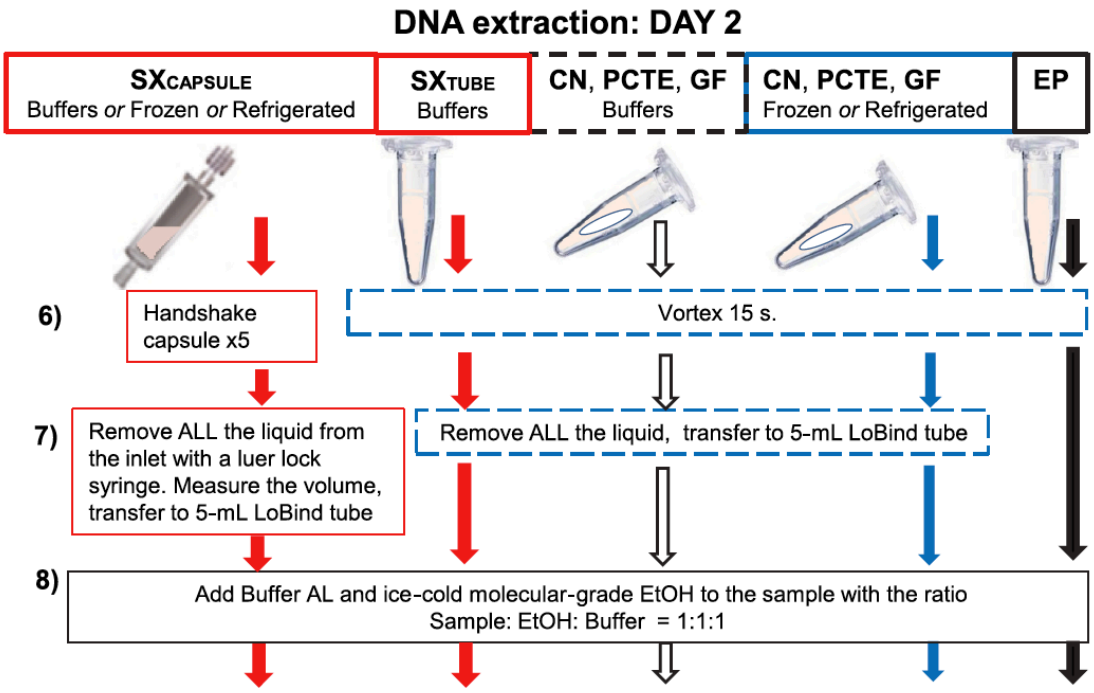


Figure. 2.1. Flow chart illustrating the modified environmental DNA (eDNA) extraction protocol. This is based on DNeasy Blood & Tissue Kit (QIAGEN, Carlsbad, CA, USA). *) Capture: SX, Sterivex-GP polyethersulfone capsule filters. Note that $SX_{CAPSULE}$ and SX_{TUBE} are treated as separate samples from step 2. CN, cellulose nitrate; PCTE, polycarbonate track-etched; GF, glass fibre filters; EP, ethanol precipitation. Storage: Frozen at -20 °C, Refrigerated are samples stored at 8–10 °C and processed within 5 h. Steps 9–26 see Appendix S1.

2.4 Results

2.4.1 Species detection

Altogether 713 qPCR samples, including controls, were analysed. No samples were discarded. Perch and pike were both detected in most of the qPCR runs from the study lake (314 of 365, Fig. 2). For both species, SX_{TUBE} showed the highest overall detection rate (95% perch and 96% pike) and EP the lowest (89% perch and 56% pike; overall difference $SX_{TUBE} \neq$ EP: Pearson $x2$ (1, n = 62) = 6.9, Fisher's exact P = 0.02).

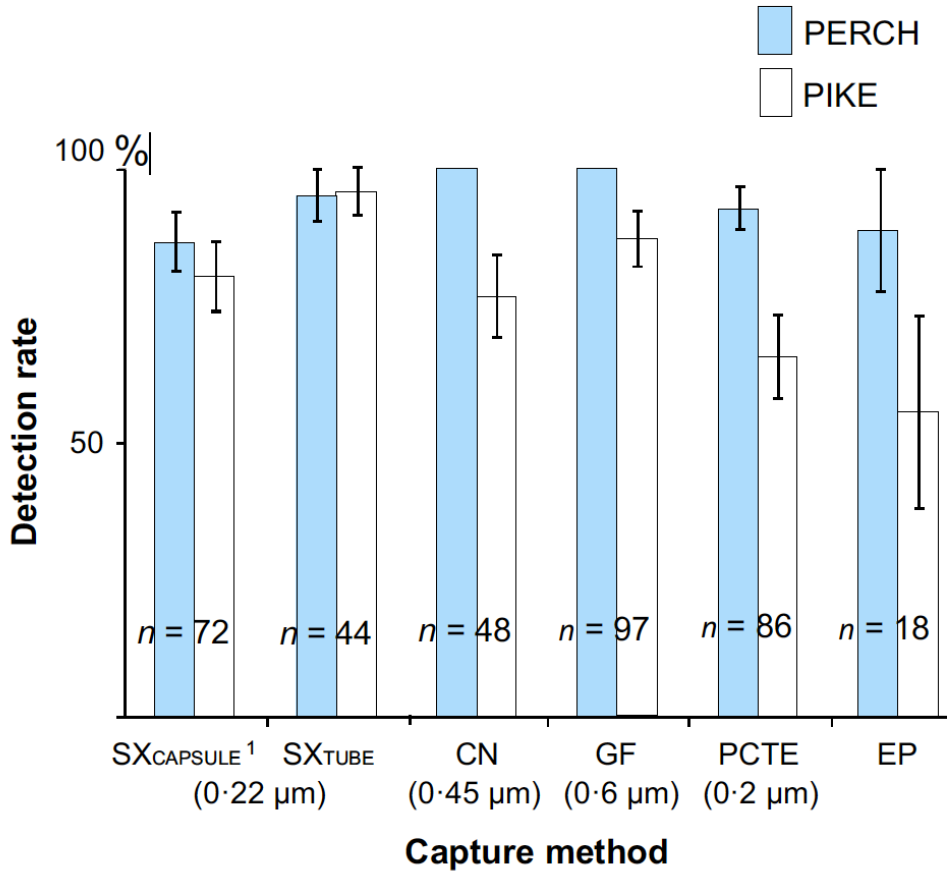


Figure. 2.2 Detection rate using quantitative PCR (qPCR; study lake).
Blue bars and clear bars show positive detections of perch and pike, respectively. Pore size of filters within parentheses. $SX_{CAPSULE}$, Sterivex, extraction within filter capsule; SX_{TUBE}, Sterivex, extraction in tube outside capsule from removed preservation buffer; CN, cellulose nitrate; PCTE, polycarbonate track-etched; GF, glass fibre; EP, ethanol precipitation. Error bars represent standard errors; *n* indicates number of trials pooling all replicates for each method and both species combined. [1]Deviating from protocol, 12 $SX_{CAPSULE}$ replicates were over-vortexed and tested mainly negative. If these 12 over-vortexed samples are omitted, the detection rate estimate for $SX_{CAPSULE}$ increases to 100% for perch and to 91% for pike.

2.4.2 Capture method

A CHAID tree multivariate predictive model was successfully generated from perch Cq-values. Capture method was the best overall predictor of Cq-values, better than storage media or storage time. In general, the lowest Cq-values were generated from $SX_{CAPSULE}$ samples in comparison with other capture methods (Fig. 2.3a). We validated the fundamental first-level outcome from this multivariate model for perch with new data in the build of a second CHAID tree, modelling pike Cq-values (Fig. 2.3b). In this second variant, capture was also the best predictor of Cq-values and $SX_{CAPSULE}$ tied with the CN and GF filters in the lowest value category. The fundamental first-level outcome of both the CHAID tree multivariate predictive models was supported in a one-by-one comparison of capture methods including both species and all treatments. Overall, $SX_{CAPSULE}$ was more efficient than the other capture methods apart from CN. $SX_{CAPSULE}$ yielded significantly higher [$eDNA_{tot}$] and lower Cq-values (Table 2).

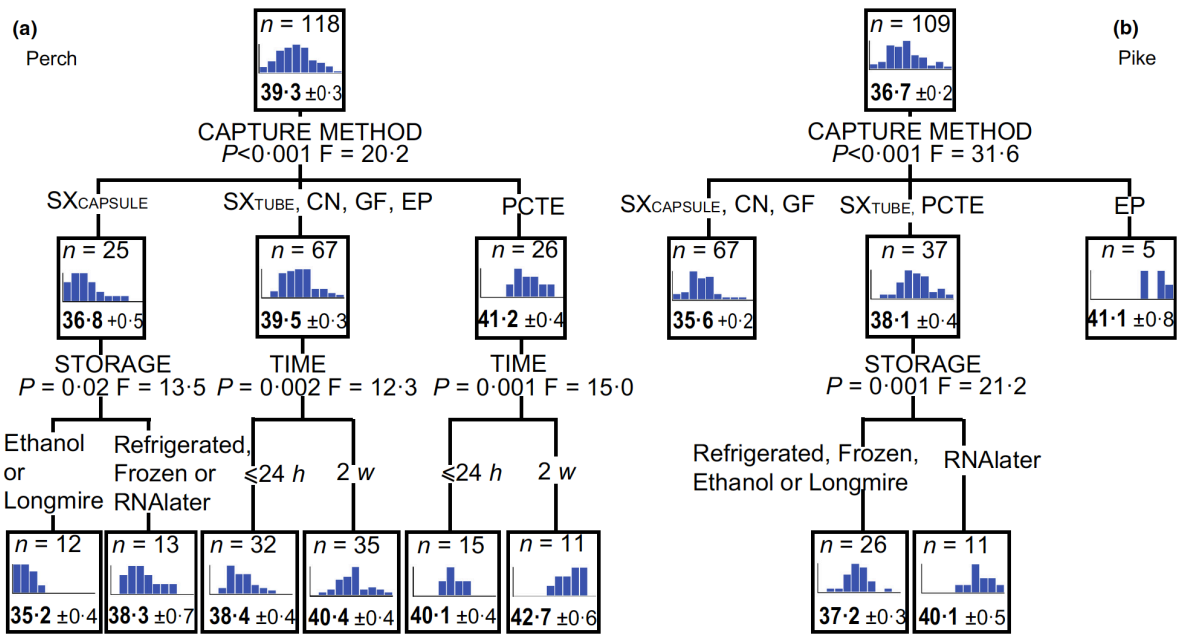


Fig. 2.3 Chi-square Automatic Interaction Detector decision trees. Relating three categorical variables (capture method, storage treatment and storage time) as model predictors for Cq-values as response target (study lake). (a) Perch. Best predictor was capture method, followed by storage time, and finally, storage treatment. (b) Pike. Best predictor was capture method followed by storage treatment. $SX_{CAPSULE}$, Sterivex, extracted within capsule; SX_{TUBE}, Sterivex, extraction in tube outside capsule; CN, cellulose nitrate; GF, glass fibre; PCTE, polycarbonate track-etched fibre; EP, ethanol precipitation; h, hours; w, weeks. Blue bar charts indicate relative size distribution of Cq-values within each category before split. Number under bar charts indicate mean Cq-value for the given category

Table 2.2 $SX_{CAPSULE}$ in comparison with other eDNA capture methods
$SX_{CAPSULE}$ comparison of Cq-values ($SX_{CAPSULE}$ comparison of [$eDNA_{tot}$]). Wilcoxon matched-pair signed-rank test of both Cq-values from qPCR and [$eDNA_{tot}$] (denoted in parentheses). Significant P-values are in bold and non-significant P-values are denoted as N.S. $SX_{CAPSULE}$, Sterivex, extracted within capsule; SX_{TUBE}, Sterivex, extraction in tube outside capsule; GF, glass fibre; PCTE, polycarbonate tracketched filter; CN, cellulose nitrate; EP, ethanol precipitation; [$eDNA_{tot}$], total eDNA concentration. *Bonferroni corrected (5 tests): a = 0.05 lowered to 0.01, a = 0.01 lowered to 0.002 and a = 0.001 lowered to 0.0002. †Due to time constraints, CN(24 h) were cancelled reducing sample size and statistical power for CN in comparison.

Capture	Pairs of *n*	P	Significance*	Z	Rank
SX_{TUBE}	33 (18)	1×10^{-5} (5×10^{-4})	*** (**)	-4.4 (-3.5)	$SX_{CAPSULE} < SX_{TUBE}$ ($>SX_{TUBE}$)
GF	50 (27)	7×10^{-3} (2×10^{-5})	* (***)	-2.7 (-4.3)	$SX_{CAPSULE}$ < GF (>GF)
PCTE	44 (27)	1×10^{-5} (6×10^{-6})	*** (***)	-4.4 (-4.5)	$SX_{CAPSULE}$ < PCTE (>PCTE)
EP	13 (9)	1×10^{-3} (8×10^{-3})	** (*)	-3.2 (-2.7)	$SX_{CAPSULE}$ < EP (>EP)
CN†	29 (15)	0.32 (0.55)	N.S. (N.S.)	-1.0 (-0.6)	

SX samples contained up to 118 ng total eDNA µL $^{-1}$ and most $SX_{CAPSULE}$ amplified before 36 cycles (Fig. 2.4). [$eDNA_{tot}$] from the fish-free control pond showed a similar pattern, being higher for CN and $SX_{CAPSULE}$ compared with GF and PCTE (Mann–Whitney U = 12, $n_1 = n_2 = 10$, Fisher's exact P = 0.003), but with no Cq-values from qPCR as target species were not present. Overall, capture method and [$eDNA_{tot}$] were fundamental predictors of Cq-values (Fig. 2.4).

2.4.3 Storage preservative

SX-specific storage results are singled out and illustrated in Fig. 2.5. SX_{TUBE} samples treated with RNAlater, a significant predictor of poorer Cq-values in the CHAID trees, were least successful. For $SX_{CAPSULE}$, preservation in ethanol or Longmire buffer improved Cq values for perch in comparison with frozen, 5 h and preservation in RNAlater (Figs 2.3a and 2.6). Also for both species pooled, these two buffers (ethanol or Longmire) in $SX_{CAPSULE}$ resulted in lower Cq-values compared with frozen or 5 h (Mann–Whitney Test *U*: 35, $n_1 = 23$, $n_2 = 15$, Z = -4.1; P = 4 x 10 $^{-5}$).

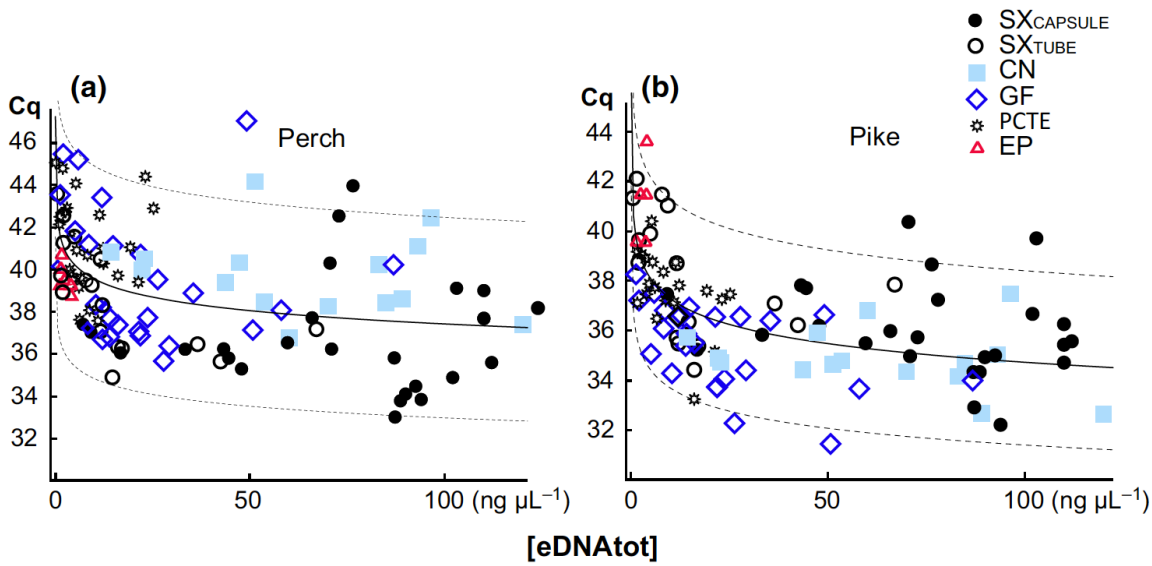


Figure 2.4 Environmental DNA (eDNA) capture methods: relationship between total eDNA concentration ([$eDNA_{tot}$]) and quantification cycles in qPCR (Cq-value) in study lake. Line represents best-fit power function where Cq decreased as a function of [$eDNA_{tot}$]. (a) Perch. Cq = 41.8 x [$eDNA_{tot}$] $^{-0.024}$; P < 0.001, R^2 = 0.23. (b) Pike: Cq = 40.0 x [$eDNA_{tot}$] $^{-0.031}$; P < 0.001, R^2 = 0.42. Dotted lines represent lower or upper limits of 95% CI for slope of regression. $SX_{CAPSULE}$, Sterivex, extracted within capsule; SX_{TUBE}, Sterivex, extracted from buffer in tube outside capsule; CN, cellulose nitrate; GF, glass fibre; PCTE, polycarbonate track-etched fibre; EP, ethanol precipitation.

2.4.4 Storage time

Storage time in the second-level outcome from the first CHAID tree was classified as a positively correlated predictor of Cq-values for all capture methods apart from SX (Fig. 2.3a). This was supported in a one-by-one comparison of capture methods including both species and 24 h to 2 weeks treatments (Table 2.3). Cq-values did not increase significantly with time using SX, but did with GF and PCTE. The mean difference between Cq-values of paired qPCR replicates run within the same day was +0.3 ± 0.2 SE. This difference increased to +1.3 ± 0.2 SE when replicates run on different days were included, indicating that freezing and thawing of eDNA once or twice between measurements decreased DNA quality [Welch's test $t(1, 68) = 7.1$, $n_1 = 20$, $n_2 = 80$, $P = 9 \times 10^{-10}$]. To avoid introducing this error, only DNA templates thawed for the first time were included when calculating average Cq-values for the samples.

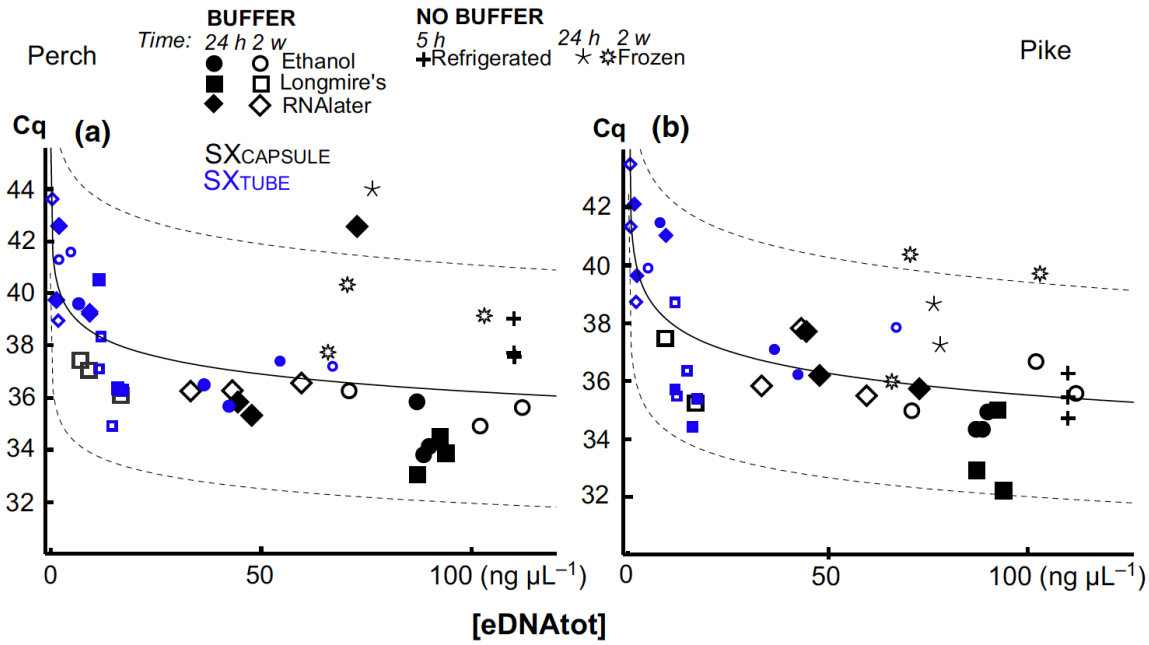


Figure 2.5. Environmental DNA (eDNA) storage treatment using SX: relationship between total eDNA concentration ([$eDNA_{tot}$]) and quantification cycles in qPCR (Cq-value) in study lake. Line represents best-fit power function of the negative correlation between Cq and [$eDNA_{tot}$]. (a) Perch: Cq = 40.9 9 [$eDNA_{tot}$] $^{-0.026}$; P < 0.001, R2 = 0.28. (b) Pike: Cq = 40.8 9 [$eDNA_{tot}$] $^{-0.030}$; P < 0.001, R2 = 0.45. Dotted lines represent lower or upper limits of 95% CI for slope of regression. Sterivex, extracted within capsule ($SX_{CAPSULE}$) and from buffer in tube outside capsule (SX_{TUBE}) shown in black and blue symbols, respectively. h, hours; w, weeks.

2.4.5 Contamination

One false-positive signal for perch was detected at 42 cycles in an EP 'no-water' negative control. Remaining negative controls for capture/storage treatments (n = 80) and negative pond water (n = 85), NTCs (n = 64) and 37/40 tissue negative controls for species specificity did not amplify. The contaminated tissue control was replaced and showed no amplification. One extraction blank came up positive in one of the seven runs, but at a very high Cq of 46.2.

Table 2.3. Effect of storage time for eDNA results with different capture methods
Paired test of Cq-values

Storage	Pairs of *n*	*P*	Significance*	Z	Rank
$SX_{CAPSULE}$	20	0.15	N.S.	-1.5	
SX_{TUBE}	16	0.18	N.S.	-1.3	
PCTE	16	0.002	**	-3.1	PCTE 24 h < PCTE 2 weeks
Glass fibre (GF)	24	0.002	**	-3.1	GF 24 h < GF2 weeks

2.5 Discussion

To our knowledge, this is the first study comparing enclosed filters (SX) with commonly used eDNA capture and storage techniques. Similarly to other capture methods, SX can be used to target a wide range of macro-organisms successfully (using PCR, qPCR or NGS; Table S1), ensuring the generality of SX for surveys of aquatic biodiversity. Specifically, SX with added preservation buffer (ethanol or Longmire's) is the optimal approach of the tested treatments in regard to [$eDNA_{tot}$] yield and detection sensitivity for target species. Other eDNA studies of macrobiota using SX (Keskin 2014; Bergman *et al*. 2016) did not apply preservation buffers. Although our study set-up was different, the lake sample results are consistent with the mesocosm experiment of Renshaw *et al*. (2015), showing that open CN filter and polyethersulfone filters (same material as SX in this study) were more effective than PCTE and GF. Additionally, we demonstrate that SX eDNA retains integrity over time, whereas eDNA from the open filters degrades significantly. These results suggest that SX eDNA is more effectively preserved, possibly due to the fact that it is considerably less handled by the user. The capsule may reduce risks of exposure to physical and biogenic stress as well as contamination, because capture, storage and extraction take place within the filter capsule. This, together with extended field usage possibilities, and higher eDNA yields, constitutes reasons to recommend enclosed filters before other capture methods.

2.5.1 Capture method

Based on our results, we reject H_0 hypothesis 1 stating that SX and commonly used techniques in our study are equally effective, because $SX_{CAPSULE}$ yields the lowest Cq-values for perch (Fig. 2.3a). However, this is only partially validated in the case of pike (Fig. 2.3b), where $SX_{CAPSULE}$, GF and CN group together for the lowest Cq-values. Overall, $SX_{CAPSULE}$ yields higher [$eDNA_{tot}$] and generates better qPCR results than other capture methods, with the exception of CN. Our CN/SX comparisons are not as extensive as the SX/GF and SX/PCTE comparisons (Table 2.2). We show that higher levels of [$eDNA_{tot}$] are related to lower Cq-values of target species DNA (R^2 = 0.23–0.45, Figs 2.4 and 2.5) and therefore suggest measurements of [$eDNA_{tot}$] for approximate indications of eDNA capture efficiency. The comparison in this study of SX_{TUBE} to $SX_{CAPSULE}$ demonstrates that utilizing both these sources of eDNA should be useful. Pooling of these in the final elution step would be advisable for gaining even higher final yields of eDNA. SX_{TUBE} exhibits the highest overall detection rate for both species (95–96%) in our study, significantly higher than EP results. Higher amounts of false negatives from EP field samples may be due to DNA retention in the falcon tubes (Gaillard & Strauss 1998) and/or to the low water volume processed (0.015 L; Deiner *et al*. 2015; Eichmiller, Miller & Sorensen 2016; Minamoto *et al*. 2016).

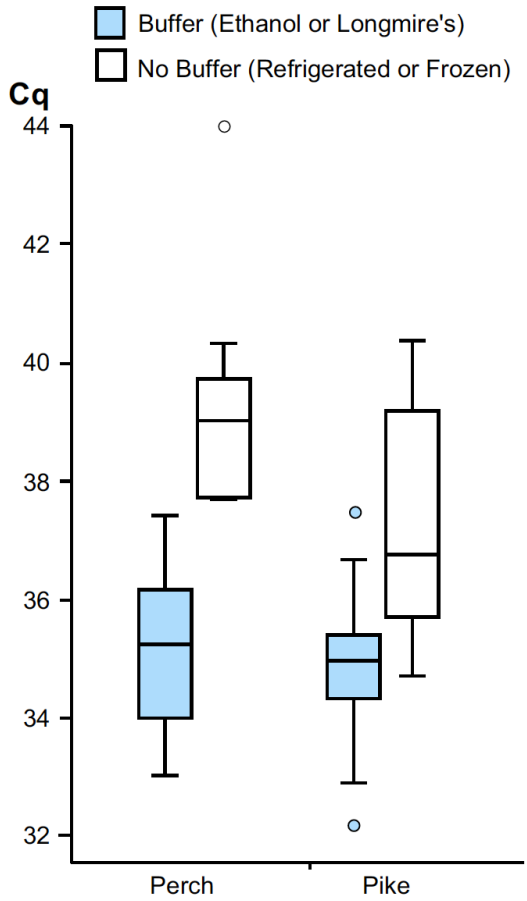


Fig. 2.6. Boxplots of Cq-values showing $SX_{CAPSULE}$ (extraction within Sterivex capsule) filter storage with and without preservation buffer (ethanol or Longmire's).

2.5.2 Storage preservative

We reject H_0 hypothesis 2a stating that preservation buffers for storage of SX do not affect qPCR amplification in comparison with extraction within 5 h or freezing at -20 °C. Two-thirds of published aqueous eDNA surveys reporting storage details apply freezing of filters as a preservation method (Table S1 and S2), while less than one-third of surveys use buffer storage. Our results indicate that addition of ethanol or Longmire's immediately after SX filtration provides the lowest Cq-values, and is significantly better than freeze storage or extraction within 5 h. Based on our results as well as the results of three previous studies (Renshaw *et al*. 2015; Wegleitner *et al*. 2015; Minamoto *et al*. 2016), we recommend addition of preservation immediately after filtration.

2.5.3 Storage time

We reject H0 hypothesis 2b that degradation of captured eDNA is the same in SX filters and the other capture techniques tested in this study. Cq-values increase significantly with storage time for GF and PCTE samples, indicating degradation of eDNA. In contrast, Cq-values for SX samples ($SX_{CAPSULE}$ or SX_{TUBE}) do not differ significantly after 2 weeks of storage at RT. We note that repeated use of the same extracted eDNA sample (eluted in TE-buffer) for qPCR on different days, entailing repeated freezing and thawing, resulted in higher Cq-values. Freeze–thaw-induced degradation and/or inhibition of DNA is previously acknowledged (e.g. Ross, Haites & Kelly 1990; Takahara, Minamoto & Doi 2015). We therefore recommend that extracted eDNA samples are divided into many aliquots immediately after extraction, in order to avoid compromising eDNA quality by repeated freezing and thawing.

2.5.4 Contamination

We cannot yet reject H_0 hypothesis 3 stating that SX leads to as many false positives as typically used methods. We only produced one false positive (EP) which is insufficient for any statistical inference. The SX approach using sealed pre-sterilized equipment until sampling, and capping filter immediately after filtration, should reduce contamination risk. The contamination variance between these capture methods remains to be tested using more observations and possibly synthetic controls (Wilson, Wozney & Smith 2016).

2.5.5 Limitations

The hand-held syringe used with SX filter units is convenient but turns into a labour-intensive bottleneck when processing many samples. This can be alleviated by switching to battery powered pumps (SterivexTM 2013). In 'algal soup' or turbid waters, 0.2 µm pore size may pose a problem as the filters clog easily and less water can be processed (Turner *et al*. 2014a).

This can be overcome by pre-filtering (Robson *et al*. 2016) and/or increasing the number of filter replicates. Future research is needed to identify optimal procedures for highly productive and/or turbid waters.

2.6 Conclusion

In conclusion, we recommend SX filters as an efficient capture method for aqueous eDNA sampling of macro-organisms. Preservation of SX in ethanol or Longmire's buffer immediately after filtration is recommended. Preserved SX capsules may be stored at RT for at least 2 weeks without significant degradation. Water samples can be quickly filtered and preserved on site requiring less equipment, easing transport. Therefore, SX capsules are logistically compatible with remote and harsh field conditions.

2.6.1 Authors' contributions

M.H., J.S., A.E and S.S.T.M conceived and designed initial experiment. All authors (except D.H.) contributed to final design and participated in 'sample collection/filtration day'. J.S. analysed data and drafted the manuscript. M.H. developed protocol for eDNA capture/extraction. J.S., M.H. and A.E. wrote the manuscript. A.E. and S.S.T.M. coordinated field experiment and contributed to extraction protocol. A.E., M.H., S.W.K., S.S.T.M., E.E.S. and M.S. extracted DNA. S.W.K. optimized qPCR protocol. S.W.K., M.H. and M.S. performed qPCR assays. All authors revised the manuscript. No conflict of interest exists.

2.6.2 Acknowledgements

We thank Philip Francis Thomsen (PFT), Ian Eirød and Peter Rask Møller for participating in 'experimental planning' and 'collection/filtration' day; PFT for providing qPCR primers/probes and collaborating on earlier SX extraction protocols; EskeWillerslev for laboratory facilities, funding and comments on earlier drafts.

2.6.3 Data accessibility

Data are deposited in the Dryad Data Repository http://dx.doi.org/10.5061/dryad.p2q4r (Spens *et al*. 2016).

2.7 References

Bergman, P.S., Schumer, G., Blankenship, S. & Campbell, E. 2016. Detection of adult green sturgeon using environmental DNA analysis. *PLoS One*, 11, e0153500.

Bustin, S.A., Benes, V., Garson, J.A. *et al*. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55, 611–622.

Chestnut, T., Anderson, C., Popa, R., Blaustein, A.R., Voytek, M., Olson, D.H. & Kirshtein, J. 2014. Heterogeneous occupancy and density estimates of the pathogenic fungus Batrachochytrium dendrobatidis in waters of North America. *PLoSOne*, 9, e106790.

Davy, C.M., Kidd, A.G. & Wilson, C.C. 2015. Development and validation of environmental DNA (eDNA) markers for detection of freshwater turtles. *PLoSOne*, 10, e0130965.

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.

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.

Eichmiller, J.J., Miller, L.M. & Sorensen, P.W. 2016. Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Molecular Ecology Resources*, 16, 56–68.

Ellison, S.L., English, C.A., Burns, M.J. &Keer, J.T. 2006. Routes to improving the reliability of low level DNA analysis using real-time PCR. *BMC Biotechnology*, 6, 33.

Ficetola, G.F., Miaud, C., Pompanon, F. & Taberlet, P. 2008. Species detection using environmental DNA from water samples. *Biology letters*, 4, 423–425.

Gaillard, C. & Strauss, F. 1998. Avoiding adsorption of DNA to polypropylene tubes and denaturation of short DNA fragments. Technical Tips Online, 3, 63–65.

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, pp.1-3.

Höss, M., Kohn, M., Paabo, S., Knauer, F. & Schroder, W. 1992. Excrement analysis by PCR. Nature, 359, 199.

IBM Corp. 2013. IBM SPSS Statistics for Windows, Version 22.0. IBM, Armonk, NY, USA.

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.

Keskin, E. 2014. Detection of invasive freshwater fish species using environmental DNA survey. *Biochemical Systematics and Ecology*, 56, 68–74.

Liang, Z. & Keeley, A. 2013. Filtration recovery of extracellular DNA from environmental water samples. Environmental Science & Technology, 47, 9324–9331.

Longmire, J.L., Maltbie, M. & Baker, R.J. (1997) Use of "lysis buffer" in DNA isolation and its implication for museum collections. Occasional Papers the Museum Texas Tech University, 163, 1–3.

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 Research, 39, 541–548.

McKee, A.M., Calhoun, D.L., Barichivich, W.J., Spear, S.F., Goldberg, C.S. & Glenn, T.C. (2015) Assessment of environmental DNA for detecting presence of imperiled aquatic amphibian species in isolated wetlands. Journal of Fish and Wildlife Management, 6, 498–510.

Minamoto, T., Naka, T., Moji, K. & Maruyama, A. (2016) Techniques for the practical collection of environmental DNA: filter selection, preservation, and extraction. Limnology, 17, 23–32.

Piaggio, A.J., Engeman, R.M., Hopken, M.W., Humphrey, J.S., Keacher, K.L., Bruce, W.E. & Avery, M.L. (2014) Detecting an elusive invasive species: a diagnostic PCR to detect Burmese python in Florida waters and an assessment of persistence of environmental DNA. Molecular Ecology Resources, 14, 374– 380.

Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R.M.&Gough, K.C. (2014) 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.

Renshaw, M.A., Olds, B.P., Jerde, C.L., McVeigh, M.M. & Lodge, D.M. (2015) The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol–chloroform–isoamyl alcohol DNA extraction. Molecular Ecology Resources, 15, 168–176.

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.

Ross, K.S., Haites, N.E. & Kelly, K.F. (1990) Repeated freezing and thawing of peripheral blood and DNA in suspension: effects on DNA yield and integrity. Journal of Medical Genetics, 27, 569–570.

Simmons, M., Tucker, A., Chadderton,W.L., Jerde, C.L. & Mahon,A.R. (2016) Active and passive environmental DNA surveillance of aquatic invasive species. Canadian Journal of Fisheries and Aquatic Sciences, 73, 76–83.

Smart, A.S., Tingley, R., Weeks, A.R., van Rooyen, A.R. & McCarthy, M.A. (2015) Environmental DNA sampling is more sensitive than a traditional survey technique for detecting an aquatic invader. Ecological applications, 25, 1944–1952.

Spens, J., Evans, A.R., Halfmaerten, D., Knudsen, S.W., Sengupta, M.E., Mak, S.S.T., Sigsgaard, E.E. & Hellstr€om, M. (2016) Data from: Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.p2q4r

Sterivex™. 2013. User Guide Sterivex™-GP Sterile Vented Filter Unit, 0.22 µm Single Use Only. EMD Millipore Corporation, Billerica, MA, USA.

Takahara, T., Minamoto, T. & Doi, H. 2015. Effects of sample processing on the detection rate of environmental DNA from the Common Carp (Cyprinus carpio). *Biological Conservation*, 183, 64–69.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., Orlando, L. & Willerslev, E. 2012. Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21, 2565–2573.

Turner, C.R., Barnes, M.A., Xu, C.C.Y., Jones, S.E., Jerde, C.L. & Lodge, D.M. 2014a. Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution*, 5, 676–684.

Turner, C.R., Miller, D.J., Coyne, K.J. & Corush, J. 2014b. Improved methods for capture, extraction, and quantitative assay of environmental DNA from Asian bigheaded carp (Hypophthalmichthys spp.). *PLoS One*, 9, e114329.

Valentini, A., Taberlet, P., Miaud, C. *et al*. 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, 25, 929–942.

Wegleitner, B., Jerde, C., Tucker, A., Chadderton, W.L. & Mahon, A. 2015. Long duration, room temperature preservation of filtered eDNA samples. *Conservation Genetics Resources*, 7, 789–791.

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.

Willerslev, E., Hansen, A.J., Binladen, J. *et al*. 2003. Diverse plant and animal genetic records from holocene and pleistocene sediments. *Science*, 300, 791–795.

Wilson, C.C., Wozney, K.M. & Smith, C.M. (2016) Recognizing false positives: synthetic oligonucleotide controls for environmental DNA surveillance. *Methods in Ecology and Evolution*, 7, 23–29.

Zhan, A.B., Hulak, M., Sylvester, F. *et al*. 2013. High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution*, 4, 558–565.

Received 29 May 2016; accepted 21 September 2016 Handling Editor: Douglas Yu

2.8 Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Fig. S1. Flow chart illustrating the different capture and storage treatments.

Appendix S1. eDNA extraction protocol.

Appendix S2. Water quality in Gentofte lake.

Table S1. Empirical field-studies targeting macrobial eDNA in aquatic ecosystems with water sampling, January 2005 to March 2015.

Table S2. Empirical field-studies targeting macrobial eDNA in aquatic ecosystems with water sampling, published after the current study was initiated in March 2015.

Table S3. Primers and probes used in this study.

Chapter 3

Universal Methods: informing field and laboratory methods for Chapter 4 and 5

3.0 Introduction

This Chapter describes the field, laboratory and bioinformatic methods used in the following Chapter 4 and Chapter 5 so that methodological explanations are not repeated across chapters. Further explanation to these descriptions are found in Chapter 4 and 5.

3.1 Sterivex Filter eDNA sampling

Filtration of eDNA was performed using the Sterivex™ Filter Units (SVGPL10RC µm, polyethersulfone, with Luer outlet, gamma irradiated, 2 L, Male Luer-Lok®). Below are the step-by-step instructions used for the isolation of eDNA through filtration using these filters.

3.1.1 Equipment

- 1 sampling bag = 1 x Sterivex filter, 1 x inlet cap, 1 x outlet cap, 2 x parafilm
- Clipboard, pencils, sampling sheet, and protocol
- Lab gloves
- Spray bottle containing 20% bleach and 80% bottled drinking water
- Spray bottle containing clean ethanol
- Unopened paper towels.
- Ice box containing frozen ice blocks (both previously sterilised with 50% bleach solution)
- If also using a storage buffer e.g. RNA Later, ethanol, or Longmire's solution
    - 2 ml per filter of buffer
    - 1,000 µl pipette
    - 1,000 µl pipette tips

3.1.2 Sampling protocol

See Figure 3.1 below.

1. Wash hands with soap and put on gloves.
2. Take a sampling bag and label both the bag and filter with permanent pen, adding plastic tape over the writing on the filter to prevent smudging during extraction.
3. Remove 60 ml syringe from sterile packaging.
4. Draw 50 ml of desired water up into the syringe.
5. Attach Sterivex filter to the syringe by gently pushing the syringe tip inside the inlet of the Sterivex, and gently twisting until the Sterivex filter is secure.

6. Push the 50 ml of water through the Sterivex filter (the water will come out of the Sterivex outlet), without applying too much pressure as this can break the filter.

7. Unscrew the Sterivex again, repeat steps 4. – 6. until completing the desired volume.

8. Draw air only into the entire syringe, and push the air through the Sterivex filter to remove remaining water droplets. The Sterivex filter must be as dry as possible.

9. If using a storage buffer, inject 2ml using the pipette and tip gently inside the Sterivex inlet. If not using storage buffer, leave inside dry.

10. Screw the Inlet Cap onto the Sterivex inlet, and the Outlet Cap onto the Sterivex outlet. Wrap both ends in parafilm

11. Put the filter back inside the labelled bag

12. Put immediately into a freezer box with frozen ice blocks inside, and transfer to -20°C freezer as soon as possible.

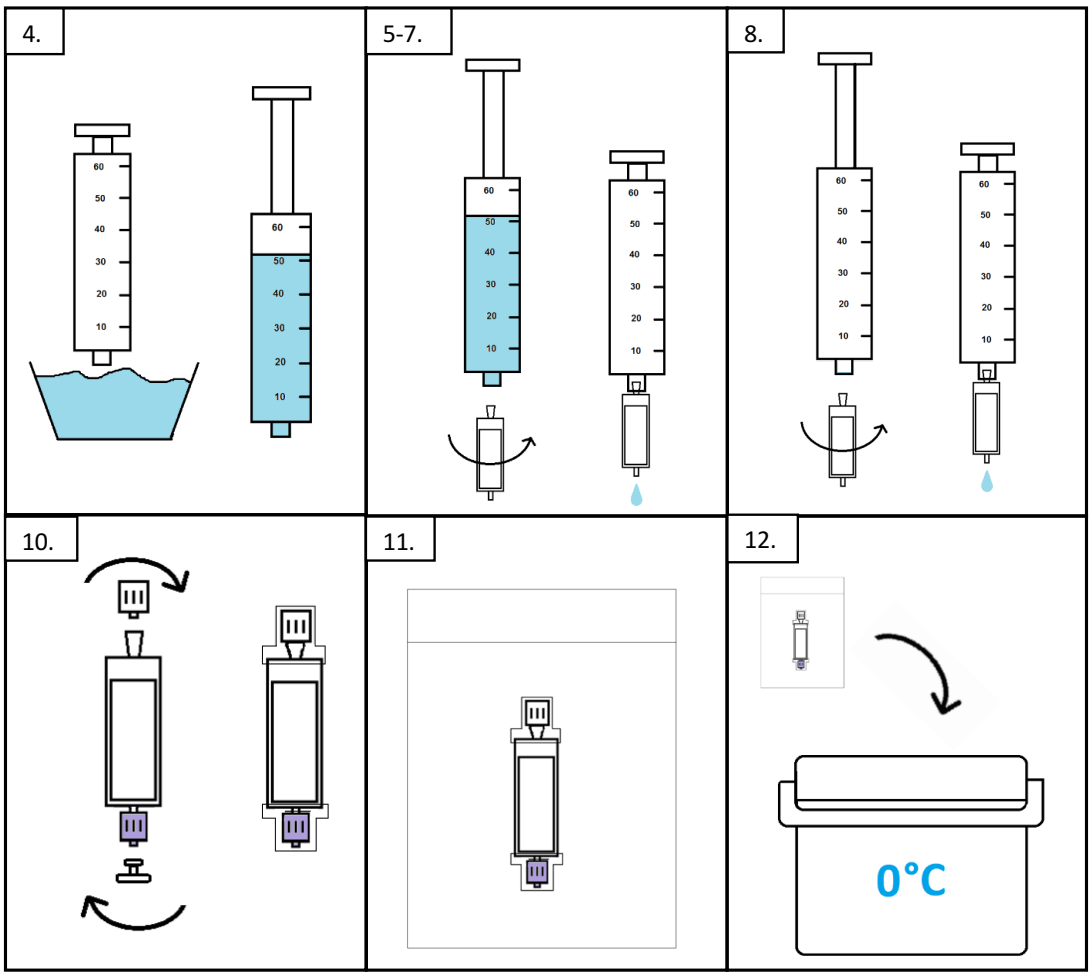


Figure 3.1 Sterivex filter sampling protocol.

3.1.3 Habitat measurements

Temperature, pH and lake depth were measured using automatic digital samplers. Turbidity was measured using a secchi disk. Nitrate (NO3) levels were measured using the Sera Nitrate-Test kit https://www.sera.de/us/product/sera-nitrate-test-no3/, and phosphate (PO4) levels were measured using the Sera Phosphate-Test kit https://www.sera.de/us/product/sera-phosphate-test-po4/.

3.2 Sterivex filter eDNA extraction

3.2.1 Equipment

- Incubation oven set to 56°C
- Rotating plate
- Centrifuge for 24 x 2 ml Eppendorf tubes
- DNA LoBind Eppendorf tubes.
- Ethanol
- Qiagen DNEasy Blood and Tissue kit
- Pipettes
- Pipette tips

3.2.2 Sterivex filter extraction protocol

Extractions were performed using the Qiagen DNEasy Blood and Tissue Kit, following the protocol designed in Chapter 2 (Spens *et al*. 2016) (also used by Minamoto *et al.* 2012, Goldberg *et al*. 2013, Pilliod *et al.* 2013, Kelly *et al.* 2014). After adding buffer ATL and Proteinase K directly inside the Sterivex filter capsules, the capsule lids were replaced and the capsules placed inside a rotating plate and secured with plastic tape to allow a maximum number of samples to be processed, and maximise security of the samples whilst rotating. The rotating plate was placed inside an incubation oven at 56°C. Labels were written directly onto the plastic housing in pen, and individually wrapped in plastic tape to prevent marks fading or being wiped off during the rotation process. For the final step of the extraction, samples were eluted in 100 µl, and due to the long and repeated final incubation, between 80-100 µl was finally available. This extraction elute was then transferred to a LoBind 1.5 ml Eppendorf tube, wrapped in parafilm, and stored in -20 °C freezers until further use.

For Chapter 5, some samples were extracted at the Indonesian Biodiversity Research Centre, as above, and some were extracted in the Geogenetics laboratory of Copenhagen University after being posted from Malaysia and Indonesia (see Chapter 5). Samples which were extracted in Copenhagen were filtered at the lake site in Indonesia or Malaysia, placed in an ice box, stored at either -4 °C in a household freezer near the lake site, or -20 °C at the

University of Science, Malaysia, or the Indonesian Biodiversity Research Centre labs in Bali. Malaysian samples were shipped on dry ice using Fedex, and Indonesian samples were injected with 2 ml of EDTA buffer (details here) and shipped using Fedex to the Natural History Museum of Denmark. As it was not possible to ship samples on dry ice from Indonesia, adding EDTA buffer was chosen to try to preserve the samples during shipment. DNA extractions performed at Copenhagen University were done in a low-quantity DNA room specifically designed for extraction, where no post-PCR processes are permitted, or movement of persons or items from post-PCR labs allowed.

3.3 Amplification of eDNA

3.3.1 Primer validation

To test the amplification success of vertebrate eDNA using the three primer pairs, preliminary samples were collected from both Chester Zoo and the Anglesey Sea Zoo in the UK. Samples were collected in sterile 1 L Gosselin™ Round HDPE Bottles, as well as sterile 15 mL tubes (Star Lab, Cat. No. E1415-0200) immediately poured into a 50-mL centrifuge tube (Star Lab Cat. No. E1450-0200) containing 33 mL laboratory grade ethanol and 2 mL sodium acetate. These samples were extracted using the Qiagen DNEasy Blood and Tissue Kit at Bangor University's Molecular Ecology and Fisheries Genetics laboratory, and then treated as all other samples were at the GeoGenetics laboratory.

3.3.2 Screening of eDNA samples

Before experimental PCR amplification, a subset of samples was first screened to assess assay response, amplification efficiency, and inhibition using qPCR. A serial dilution of the original template was created using the dilution factors; 1:1, 1:2, 1:10 and 1:20, with qPCR performed on a qPCR machine at the GeoGenetics laboratory in Copenhagen University. This approach has been used in other metabarcoding studies (Berry *et al*. 2017). Where DNA extracts were amplified, the DNA dilution with the highest concentration of uninhibited amplification (determined by qPCR C_T values and if different amplification curves crossed over one another) was selected for subsequent metabarcoding using tagged primers (primer indexes). This type of optimisation of template DNA has been shown to improve sensitivity, reproducibility and quality of metabarcoding data (Murray, Coghlan, & Bunce, 2015).

3.3.3 PCR

Table 3.1 Group-specific mitochondrial 12S, 16S and COI primers.

Name	Sequence (5' –3')	Annealing Temperature	Reference
teleo_F	ACACCGCCCGTCACTCT	55 °C	Valentini *et al*. 2016
teleo_R	CTTCCGGTACACTTACCATG		Valentini *et al*. 2016
16Smaml	CGGTTGGGGTGACCTCGGA	59°C	Taylor *et al*. 1996
16Smam2	GCTGTTATCCCTAGGGTAACT		Taylor *et al*. 1996
jgHCO2198	TAIACYTCIGGRTGICCRAARAAYCA	52°C	Geller *et al*. 2013
mlCOIintF	GGWACWGGWTGAACWGTWTAYCCYCC		Leray *et al*. 2013

Table 3.2. Examples of studies which used the 12S, 16S and COI primers used herein.

Primer	Reference
12S (Valentini *et al*. 2016)	Hänfling, B. *et al*., 2016
12S (Valentini *et al*. 2016)	Sigsgaard, *et al*. 2017
12S (Valentini *et al*. 2016)	Thomsen *et al*. 2016
16S (Taylor *et al*. 1996)	Schnell *et al*. 2012
16S (Taylor *et al*. 1996)	Cannon *et al*. 2016
16S (Taylor *et al*. 1996)	Klymus *et al*. 2017
COI (Leray *et al*. 2013)	Kelly *et al*. 2014
COI (Leray *et al*. 2013)	Leray *et al*. 2015

Amplification and further molecular work was performed in the GeoGenetics laboratory, in three separate laboratory rooms. Room 1 for pre-PCR (no-DNA, only reagents permitted, Room 2 for pre-PCR (DNA is permitted), and Room 3 for PCR / post-PCR work. No movement of persons or items from Room 2 to Room 1, or from Room 3 to Room 2 or 1 is allowed, and fresh clothes must be worn when entering Room 1 or Room 2. Once a PCR master mix was made in Room 1, DNA was added in Room 2 where no post-PCR processes are permitted. All work was performed in a flow-hood wherever possible. Three marker genes were utilised to maximise taxonomic coverage, this multi-gene approach reduces taxonomic bias and increases taxonomic coverage (Alberdi *et al.* 2017; Stat *et al*. 2017). These were: 12S rRNA targeting teleost fish (Valentini *et al.* 2016), 16S targeting mammals (Taylor *et al.* 1996), and COI targeting all metazoa (Leray *et al.* 2013). These primers were selected for their success in previous eDNA metabarcoding studies.

All primers were individually labelled with a unique oligonucleotide 'primer index' sequence (see General Introduction), with a number of unique tag combinations (for 12S $n = 32$, for 16S $n = 59$ and for COI $n = 60$) (see Appendix 5 for details). For the 12S primer set, tags were designed using the OligoTag program (Coissac, 2012) and consisted of six nucleotides with a distance of least three bases (from Thomsen *et al*. 2015). Two or three random bases (NNN or NN) (De Barba *et al*. 2014) were attached to the end of the primer index sequence to increase complexity in the final pooled sample. Each PCR reaction was individually amplified using a matching forward and reverse tag, i.e. a 'twin tag' (the same primer index for F and R) approach (e.g. Tag1-Tag1, Tag2-Tag2, Tag3-Tag3… etc) so that each amplicon is double tagged with matching tags (see Figure 3.2), allowing increased confidence in assigning a sequence to a sample through the removal of non-twin primer index combinations (e.g. Tag1-Tag2, Tag1-Tag3, Tag2-Tag3… etc) which may arise due to tag jumping (Schnell *et al*. 2015).

Forward and reverse primers with primer indexes (tags) attached were diluted to 10 mM concentrations, and then matching tag combinations combined into seven out of eight tubes in a PCR strip in chronological order, so that the final concentration of each forward or reverse primer was 5 mM. This was done for ease of pipetting using a multi-pipette, by which the first seven wells contained a forward and reverse primer mix with matching tags, and the final eighth well was left empty. This was done so that the final well in a PCR strip was used for a negative control with either an untagged primer, or a primer with a unique tag used for all PCR negatives. To avoid primer-index (tag) related bias in amplification (O'Donnell *et al*. 2016), primer-index combinations were rotated along different samples for each PCR replicate. For each sample, 3 x replicates were performed, and when 1/3 replicates did not show amplification on a gel, the PCR was repeated, and this repeated PCR sample used. All PCR reactions were performed in 25 µl volumes of 2 µl of DNA, 2µl of forward primer (diluted to 10mM) 2 µl of reverse primer (diluted to 10um) 2.5 µl 10 x PCR Gold Buffer (Applied Biosystems Life Technologies, no MgCl2), 2.5µl MgCl (Applied Biosystems Life Technologies 25mM), 0.5 µl DNTP (Gene ON, dNTP mix), 0.2 µl AmpliTaq Gold (Applied Biosystems Life Technologies, 5U/µL), 1 µl BSA and 12.30 µl water (molecular grade). PCR conditions were as follows: 95°C for 5 minutes, then 35 cycles of 95°C for 12 seconds, x°C for 30 seconds, 70°C for 25 seconds, followed by 70°C for 7 minutes, 4°C hold. For 16S x°C = 59°C, for 12S x°C = 55°C and for COI, x°C = 52°C.

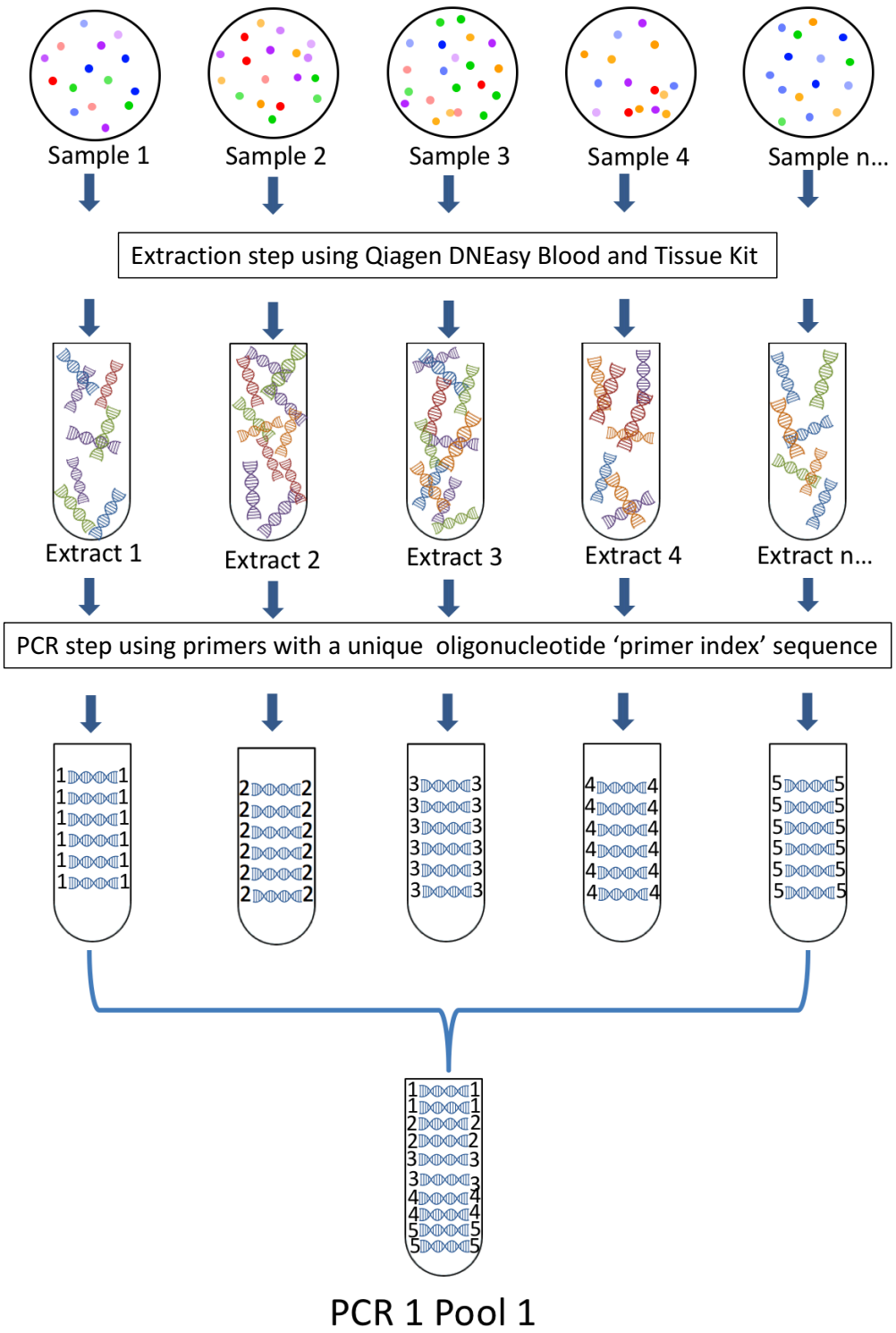


Figure 3.2 Metabarcoding set up from sample to first PCR pool. The top line of circles indicates the mix of eDNA molecules per sample which may come from different taxonomic groups, e.g. red = mammal, blue = fish, green = plant. The second line indicates the eDNA extracts created using the Qiagen DNEasy Blood and Tissue kit, containing a mix of eDNA from different taxonomic groups. The third line indicates the amplified PCR product from one differently tagged primer e.g. 12S amplifying mostly fish DNA. The final tube indicates a library pool, consisting of five differently tagged PCR samples using the same primer.

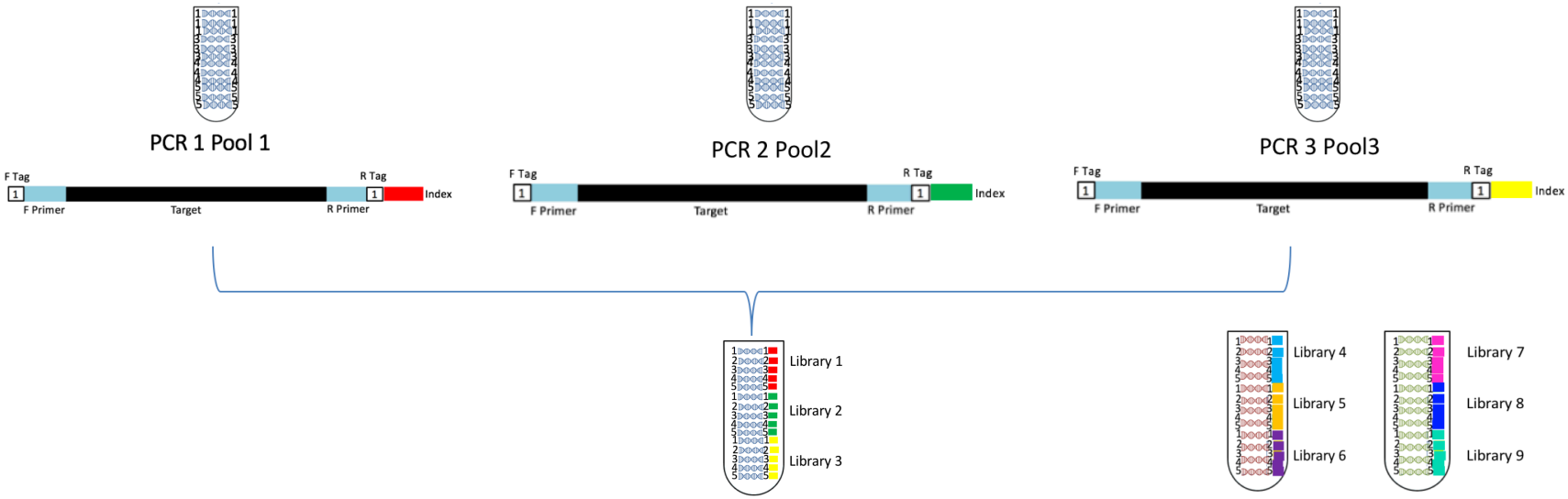


Figure 3.3 Metabarcoding set up from PCR pool to library pool. The top line shows PCR1 Pool1, PCR2 Pool2 and PCR Pool3 from the previous step (Figure 3.2) which then undergoes a second PCR to add a unique oligonucleotide index, so that the PCR pools can be combined into one library pool. For example, the centre library pool shows 3 x PCR replicates using e.g. 12S primers targeting fish, with individually tagged samples, combined into PCR pools which are also uniquely labelled with an index. Other libraries (for 16S targeting mammals, red, and COI targeting metazoan, green).

3.4 Sequencing of eDNA

3.4.1 Library Building and Sequencing

PCR products from all wells were verified on 2% agarose gels stained with GelRed™, with conditions as follows: either 2% agarose gel (for 12S and 16S), or 1.8% gel (for COI), using 120V, 400mA 120 for 40-45 minutes. The resulting images were used to assess amplified PCR product band strength, and categorise the bands by eye into four categories from which a commensurate volume was taken according to relative concentration; 1 = strong = 5 µl, 2 = medium = 7.5 µl, 3 = weak = 10 µl and 4 = no band = 12.5 µl. PCR products were then pooled for the first step of the library build protocol by combining one PCR replicate of the different samples, so that the same tag combination appeared only once per pool. The number of pools per primer pair varied according to the number of primer-indexes (tag) combinations available (see Appendix 5). Fragment size and concentration of libraries were verified on an Agilent 2100 Bioanalyzer. Library building was performed using the NEBNext® DNA Library Prep Master Mix Set for 454, using a modified NEBNext protocol combined with TruSeq indexes. Libraries were subsequently pooled in equimolar concentrations and sequenced on the Illumina MiSeq platform (½ flow cell) using 150 bp paired-end sequencing for the 12S and 16S primers, and 250 bp paired-end sequencing for the COI primers at the Danish National Sequencing Centre. To improve any low-diversity samples, a 15% spike-in of PhiX (PhiX Control v3 Library commonly referred to as PhiX, FC-110-3001, derived from the small, well characterized bacteriophage PhiX genome) was incorporated into each sequencing run to increase DNA complexity, known to improve DNA sequencing success.

3.5 Bioinformatic Analysis

Stringent sequence and taxon filtering parameters were employed with the aim of generating a high-confidence data set, removing false positives and correctly classifying true positives. False positives may have arisen through low-quality or spurious reads, low-confidence annotations, or spurious annotations. Bioinformatic analyses were implemented using a custom script, ran on Mac OS X using `python/v2.7.12`. The script used command line tools combined with various software in a pipeline shown in Figure 3.4. The summary, and details of the bioinformatic pipeline is explained below.

3.5.1 Summary of bioinformatic pipeline

1. Transfer raw reads
2. QC analysis of reads for via FastQC report
3. Trim adapters, quality check and merge paired reads
4. FastQC merged reads to create a FastQC report
5. Sort reads by tags and primers within pools
6. Confirm tag combinations on sequences within each pool
7. Filter sequences across PCR replicates
8. Check PCR replicates and positive and negative controls
9. Cluster merged reads into Operational Taxonomic Units (OTUs)
10. Create OTU table
11. Blast OTUs and open in Megan to assign taxonomy

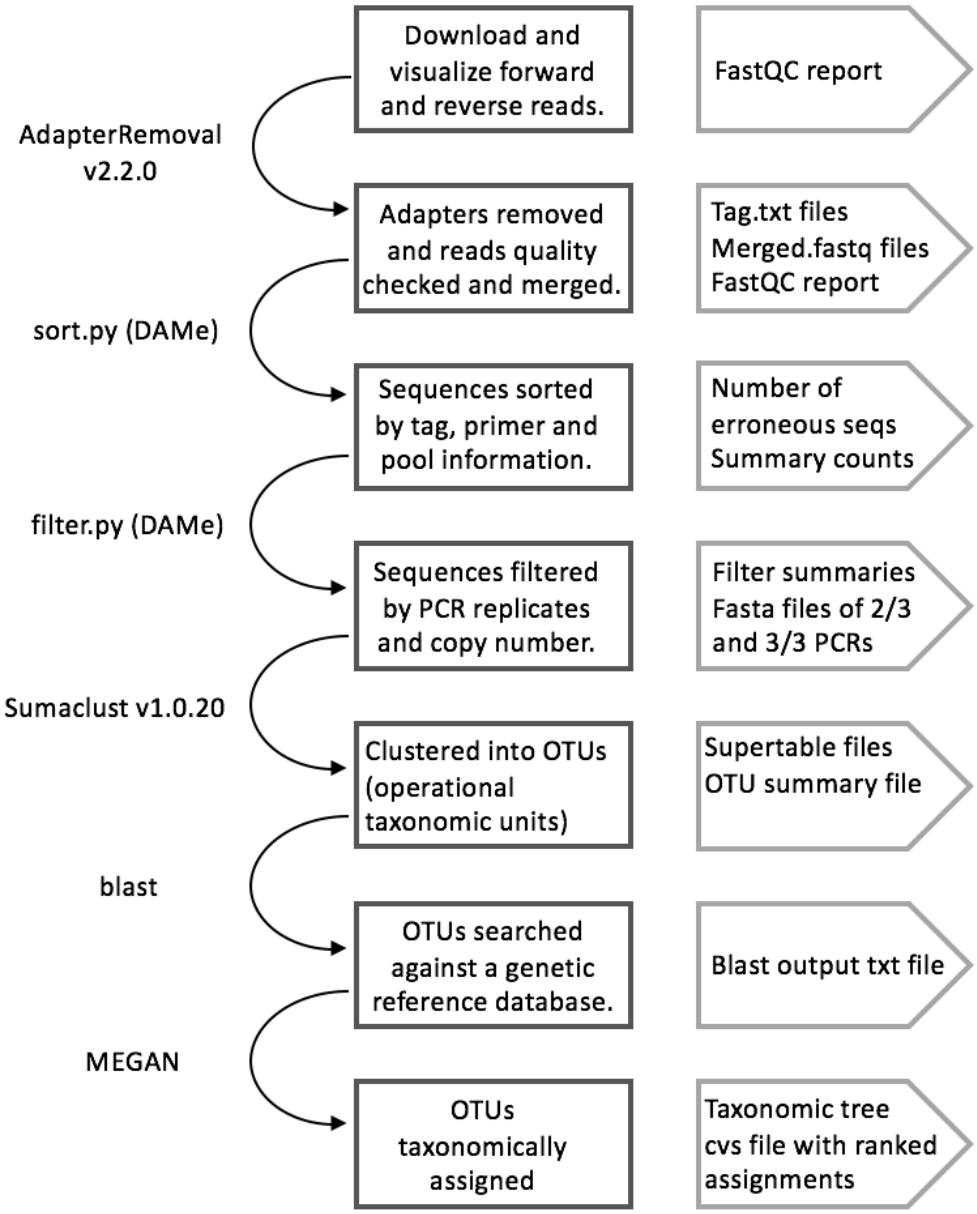


Figure 3.4 Summary of bioinformatic steps.

3.5.2 Detailed bioinformatic pipeline

Reads were transferred from a download available from the National High-Throughput Sequencing Centre, Denmark http://seqcenter.ku.dk/ to the local server @hpc.ku.dk and stored in folders referred to as 'pools' based upon individual PCR replicates grouped into different libraries. High throughput sequence quality control (QC) analysis was performed on raw reads using `fastqc/v0.11.5` to create a quality control (FastQC) report, to identify problems and assess general read information. Remnant adapter sequences were removed, and paired reads were quality checked and merged using `AdapterRemoval/v2.2.2` (Schubert

et al. 2016). Reads shorter than 50 bp were discarded following trimming. Alignments were considered where up to 5 nucleotides were missing from the 5' termini. Ambiguous bases (N) were trimmed at the 5'/3' termini. Bases were trimmed at the 5'/3' termini with quality scores <= a minimum quality PHRED value of 28, encoded upon a quality base of 33. Paired end read alignments of a minimum alignment length of 50 or more bases were collapsed, combined into a single consensus sequence representing the complete insert, and written to either basename.collapsed or basename.collapsed.truncated (if trimmed due to low-quality bases following collapse). These two file types were then merged into one fastq file, and a FastQC report created as above. Merged fastq files (amplicon sequences) were then sorted by tags and primers within pools using the program `DAMe/v0.9` (Zepeda Mendoza, *et al*., (2015) and its python script `sort.py`. Sequences were then filtered across PCR replicates using the `filter.py` script in `DAMe`, based upon the number of sequence copies found in negative controls. Sequences were only retained if they occurred in at least two out of three PCR replicates. Short reads (such as primer dimers) were filtered out by selecting a minimum sequence length (`-l`) based on expected amplicon size; for COI data, `-l` = 300, for 16S data, `-l` = 80, for 12S data `-l` =50. Abundance filtering was employed for each data set, for the COI data, sequences were retained with a minimum of 50 copies were retained, and for 12S and 16S data, sequences with a minimum of 20 copies. Using the `DAMe` python script `plotLengthFreqMetrics_perSample.py`, a graph of read counts categorised into fragment length was plotted from which the minimum and maximum length to trim sequences was decided upon. Filtered reads now within a `FilteredReads.fna` file were then converted into a file format accepted by `Usearch` or `sumaclust` (Boyer *et al*. 2014) using the `DAMe` python script `convertToUSearch.py`, using an `–lmin` (minimum length) and `–lmax` (maximum length) of 300 and 300 bp for COI data, 80 and 120 bp for 16S data, and 60-120 bp for 12S data based upon the `SequenceLengthDistribution.pdf` file. Clustering of amplicon sequences into OTUs was then performed using `sumaclust/v1.0.20` (Boyer *et al*. 2014). An identity score of 0.97 (i.e. an identity of 97%) was chosen for each dataset based upon comparisons of OTU clustering settings (Table 3.3 below). As using the `–e` or `R` arguments did not cause a large change in OTU number, these arguments were left out of the final command for clustering. It was observed (see Figure 3.5) that there was an 'inflection' where OTU number increased at a higher rate between 98-99% clustering for both 16S and 12S data, (although OTU number from COI data increased more steadily).

Table 3.3 OTU cluster testing for each marker.

Marker	Min copy number	Clustering identity	No. of OTUs	No. of OTUs with -e	No. of OTUs with R 0.9	No. of OTUs with R 0.95
COI	2	96%	4280	4280	4346	4316
COI	2	97%	4601	4601	4670	4638
COI	2	98%	5044	5044	5116	5084
COI	2	99%	5877	5877	5950	5919
COI	20	96%	1031	1031	1044	1039
COI	20	97%	1073	1073	1088	1082
COI	20	98%	1116	1116	1132	1124
COI	20	99%	1233	1233	1252	1240
COI	50	96%	583	583	589	588
COI	50	97%	599	599	606	605
COI	50	98%	616	616	622	621
COI	50	99%	658	658	665	662
16S	2	96%	146	146	148	147
16S	2	97%	174	174	177	176
16S	2	98%	513	513	518	516
16S	2	99%	2901	2901	2900	2900
16S	20	96%	86	86	86	86
16S	20	97%	88	88	88	88
16S	20	98%	185	185	185	185
16S	20	99%	609	609	609	609
12S	2	96%	214	214	216	216
12S	2	97%	849	849	853	853
12S	2	98%	958	958	962	962
12S	2	99%	6728	6728	6728	6728
12S	20	96%	123	123	123	123
12S	20	97%	151	151	151	151
12S	20	98%	155	155	155	155
12S	20	99%	1714	1714	1714	1714

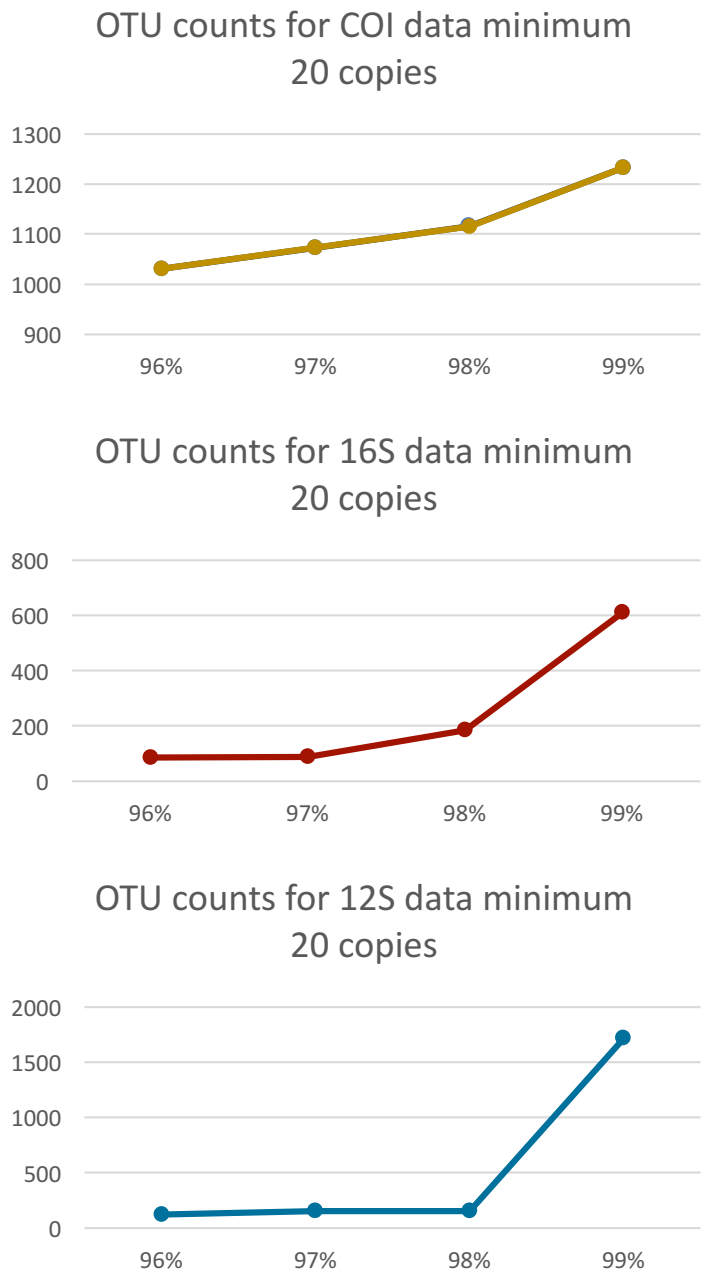


Figure 3.5 OTU counts for each marker standardised to 20 copies. COI (top), 16S (middle) and 12S (bottom) when changing the OTU cluster identity setting. Note the inflection beginning around 97%, which led to this clustering level being selected. Although 50 copies were retained for COI data, only 20 were used to create this graphical comparison for consistency.

For 16S and 12S data, only OTUs with a read count of more than 20 were retained, and for COI data, only OTUs with a read count of more than 50 were retained. Read counts were normalised using the python script `tabulateSumaclust.py` within `DAMe` using the `-s (--scale)` argument which sets the number of reads to scale each sample to. The blast input file created was then imported into MEGAN 6 (community edition) used for taxonomic assignments using the default settings, which was then linked to the OTU tables. Once OTU tables were created and populated with the taxonomic information from MEGAN, each sequence was individually verified by running a BLAST search on the NCBI database using megablast, and the output assessed for query cover, identity, and the consistency of sequences in the sequential hits. Since data is exchanged daily between EMBL-Bank and NCBI Genbank (NCBI 2017), it is assumed that there should be no major differences when using either of these databases.

3.5.3 NCBI and BLAST

Each species name given for the species assignment from BLAST was double checked by a Google search for that species, to confirm that the name was correct. Fish species were confirmed using fishbase.com, which has the most current nomenclature, and includes all synonyms.

Table 3.4 BLAST Identity accepted for each taxonomic level assignment

Taxonomic level	BLAST Identity
Species	≥ 99
Genus	95 - 98
Family	90 – 94
Order	80 – 89
Class	70 – 79
Phylum	60 – 69
Domain	≤ 59

Any sequences with a BLAST Query Cover of less than 55 % were removed from the analysis due to the likelihood of the sequence being a chimera or sequencing artefact. Only an Identity of 99-100% with no other matches to other species with the same match quality were accepted as species level assignments from BLAST, otherwise the OTU was assigned to genus level. For example, OTU20 matched with 100% Query Cover and 100% Identity to the top six hits of four different species *Sarotherodon galilaeus*, *Sarotherodon melanotheron, Oreochromis niloticus* and *Oreochromis aureu,* and so was assigned to family level.

3.6 References

Alberdi, A., Aizpurua, O., Gilbert, M.T.P. and Bohmann, K., 2017. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*.

De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, *et al*. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Mol Ecol Resour*. 2014;14: 306–323. pmid:24128180

Berry, T.E., Osterrieder, S.K., Murray, D.C., Coghlan, M.L., Richardson, A.J., Grealy, A.K., Stat, M., Bejder, L. and Bunce, M., 2017. DNA metabarcoding for diet analysis and biodiversity: A case study using the endangered Australian sea lion (Neophoca cinerea). Ecology and evolution, 7(14), pp.5435-5453.

Cannon, M.V., Hester, J., Shalkhauser, A., Chan, E.R., Logue, K., Small, S.T. and Serre, D., 2016. In silico assessment of primers for eDNA studies using PrimerTree and application to characterize the biodiversity surrounding the Cuyahoga River. *Scientific reports*, 6, p.22908.

Coissac E. OligoTag: A Program for Designing Sets of Tags for Next-Generation Sequencing of Multiplexed Samples. In: Pompanon F, Bonin A, editors. Data Production and Analysis in *Population Genomics*. *Humana Press;* 2012. pp. 13–31. Accessed at: http://dx.doi.org/10.1007/978-1-61779-870-2_2

Goldberg, C.S. *et al.*, 2013. Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshwater Science*, 32(3).

Kelly, R.P., Port, J.A., Yamahara, K.M. and Crowder, L.B., 2014b. Using environmental DNA to census marine fishes in a large mesocosm. *PLoS ONE*, *9*(1), p.e86175.

Klymus, K.E., Richter, C.A., Thompson, N. and Hinck, J.E., 2017. Metabarcoding of Environmental DNA Samples to Explore the Use of Uranium Mine Containment Ponds as a Water Source for Wildlife. Diversity, 9(4), p.54.

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

Leray, M. and Knowlton, N., 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, 112(7), pp.2076-2081.

Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M.N. and Kawabata, Z.I., 2012. Surveillance of fish species composition using environmental DNA. *Limnology*, *13*(2), pp.193-197.

Murray, D.C., Coghlan, M.L. and Bunce, M., 2015. From benchtop to desktop: important considerations when designing amplicon sequencing workflows. PLoS One, 10(4), p.e0124671.

NCBI (2017) How to submit data to GenBank. GenBank. https://www.ncbi.nlm.nih.gov/genbank/submit/.

O'Donnell, J.L., Kelly, R.P., Lowell, N.C. and Port, J.A., 2016. Indexed PCR primers induce template-specific bias in large-scale DNA sequencing studies. PloS one, 11(3), p.e0148698.

Pilliod, D.S. *et al.*, 2013. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, 70(8).

Schnell, I.B., Thomsen, P.F., Wilkinson, N., Rasmussen, M., Jensen, L.R., Willerslev, E., Bertelsen, M.F. and Gilbert, M.T.P., 2012. Screening mammal biodiversity using DNA from leeches. *Current biology*, *22*(8), pp.R262-R263.

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(6), pp.1289–1303.

Schubert, M., Lindgreen, S. and Orlando, L., 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC research notes*, 9(1), p.88.

Stat, M., Huggett, M.J., Bernasconi, R., DiBattista, J.D., Berry, T.E., Newman, S.J., Harvey, E.S. and Bunce, M., 2017. Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, 7(1), p.12240.

Taylor, P.G., 1996. Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. *Molecular biology and evolution*, 13(1), pp.283-285.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F. and Gaboriaud, C., 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, 25(4), pp.929-942.

Zepeda Mendoza, M.L., Sicheritz-Pontén, T. and Gilbert, M.T.P., 2015. Environmental genes and genomes: understanding the differences and challenges in the approaches and software for their analyses. *Briefings in bioinformatics*, 16(5), pp.745-758.

Chapter 4

The distribution of eDNA within the Indonesian lake, Danau Tamblingan: recommendations for eDNA sampling of tropical lentic habitats

4.1 Abstract

The spatial distribution of eDNA within a lacustrine environment is likely dependent on a variety of factors including degradation time, lake hydrology, animal behaviour and environmental conditions. When sampling eDNA from lacustrine habitats, it is unclear how many samples should be collected, and how far apart they should be collected to encompass the extant biodiversity. In this chapter, I investigate how the collection of aquatic eDNA from different spatial sampling points within the same lake has an effect on the biodiversity information generated through metabarcoding sequenced using the Illumina MiSeq. Nine points were sampled with three filter replicates each, at regular spatial intervals across the surface of a Balinese caldera lake (Lake Tamblingan), and ten different depth points from 0 – 18 m deep. Sterivex filters were used to filter water on site and capture eDNA, which was then amplified using three mitochondrial markers (12S, 16S and COI) to maximise the generation of biodiversity information. Fish and mammal species detected were verified by records in previous studies. Different taxonomic community composition and OTU richness was generated from different sites 500 m apart, and from different depth points only 2 m apart. This variability highlights the need for aquatic eDNA studies of standing waters to employ a sampling technique that is as spatially thorough as possible if the aim is to detect total biodiversity. However, further work testing points at more regular intervals, and storing filters in a buffer could increase the taxonomic information generated and give a clearer picture of how eDNA is spatially distributed within a tropical lake.

4.2 Introduction

4.2.1 Current understanding of eDNA

Environmental DNA (eDNA), defined by Thomsen and Willerslev (2015) as 'genetic material obtained directly from environmental samples (soil, sediment, water etc.) without any obvious signs of biological source material' has become a hot topic in the world of molecular ecology and wildlife biology, highlighted in many recent reviews from the last five years (Lodge *et al*. 2012; Yoccoz, 2012a; Taberlet *et al*. 2012a; b; Rees *et al*. 2014; Bohmann *et al*. 2014; Rees *et al*. 2015; Pedersen *et al*. 2015; Lawson Handley, 2015; Thomsen and Willerslev, 2016; Deiner *et al*. 2017b; Evans *et al*. 2017c; Hansen *et al*. 2018; Cristescu and Hebert, 2018). Within this recent surge, the most common application of environmental DNA sampling for macrobial life has been from aquatic habitats, with implications for the monitoring of aquatic wildlife for ecosystem assessment, conservation management, and

tracking of invasive species. Freshwater studies have covered a range of environments, from water bodies as vast as the Great Lakes of the USA (Jerde *et al*. 2011), to microcosms as small as the water collected in bromeliad plants of Tobago (Torresdal *et al*. 2017).

The field of eDNA research and its use in biodiversity monitoring is still in its infancy, thus, the majority of studies have mainly focused on proof of concept, and as of yet, there are few universal, standardised protocols for optimal sampling of aquatic eDNA from specific environments in the wild. The number of replicates and position of sampling points within a habitat will of course (as with traditional sampling methods) yield varying results. Results are dependent upon the probability of detecting the target taxon, which generally increases with closer spatial and temporal proximity of point-of-sampling to the target, and thus the availability of eDNA particles. Therefore, a sampling strategy that maximises the detectability of a target species or a target group must be employed as far as logistics and resources will allow. It is unclear precisely how homogeneous the distribution of eDNA from different organisms are within a water body, and exactly how this may vary across lentic and lotic systems, and warmer and colder climates. If eDNA monitoring is to be adopted by conservation managers, environmental consultants or ecotoxicologists, for example, then the ecology of eDNA (Barnes and Turner, 2016) needs to be further understood to inform best practise approaches to field sampling design.

4.2.2 What approaches are currently used for aquatic eDNA sampling?

As early eDNA studies focussed on demonstrating the concept of connecting eDNA with species identification, sampling strategies were rarely fully described, sometimes with limited information such as "samples were obtained from the river" (Martellini *et al*. 2005). There are now a limited number of official protocols for sampling of macrobial eDNA from aquatic habitats for wildlife biology and biodiversity monitoring. The United States Department of Agriculture has, in collaboration with the National Genomics Centre for Wildlife and Fish Conservation and the Forest Service, published a protocol for collecting eDNA samples from streams for fish detection, including: kit; procedures for avoiding contamination; choice of sampling location; collection of control samples; and best-practise for storing the eDNA filter (Carim *et al*. 2016). Other official protocols include how to sample pond eDNA for the detection of Great Crested Newts in the UK (Williams, 2013) and how to filter water to capture eDNA from aquatic organisms in the U.S.A. (Laramie *et al*. 2015).

For lotic systems such as rivers and streams, initial proof-of-concept studies focused on simple sampling strategies targeting areas where there was *a priori* knowledge of approximate presence or abundance of individuals against which to compare eDNA concentrations e.g. Thomsen *et al*. (2012a). When comparing eDNA sampling with traditional methods, water samples were mostly only collected at the same point of conventional sampling methods to compare the two, e.g. when comparing fyke nets to eDNA sampling for fish surveys (Shaw *et al*. 2016). Some river studies have simply collected a single surface water sample from the edge of the river at a few locations (Deiner and Altermatt, 2014; Fukumoto *et al*. 2015; Laramie *et al*. 2015; Deiner *et al*. 2016), whilst others have collected at least three samples per location, and used multiple locations (Goldberg *et al*. 2013; Pfleger *et al*. 2016; de Ventura *et al*. 2017), sometimes using a transect approach consisting of the left side, centre and right side of the river (Goldberg *et al*. 2013). Other studies report full information on sample location coordinates, time, water depth and water temperature from more than 100 samples along a river network, collected in triplicates (Pfleger *et al*. 2016).

For lentic systems such as ponds and lakes, early studies again focused on proof-of-concept, and so collected few samples with a basic approach of 3 x 15 mL samples per pond (Ficetola *et al*. 2008; Dejean *et al*. 2012; Thomsen *et al*. 2012a). Later protocols collected considerably more samples from around the pond (20 x 15 mL) (Williams, 2013), while other studies increased the sampling volume (20 x 40 mL) (Tréguier *et al*. 2014). Techniques have since generally moved from ethanol precipitation of low volume samples, to filtration of either around 1 L (Takahara *et al*. 2013; Fujiwara *et al*. 2016; Davison e*t al*. 2017), 2 L (Gingera *et al*. 2017) or 2.5 L of water (Larson *et al*. 2017) from one, or a few points per pond or lake.

By combining aquatic eDNA filtering with metabarcoding approaches and next-generation sequencing, many species can be detected at once and a rough estimate of relative abundance could be generated through observing sequencing read counts per OTU (Operational Taxonomic Unit), or exact amplicon sequence variants (see Glossary) (Clarke *et al.* 2017; Callahan *et al*. 2017). Hänfling *et al*. (2016) combined eDNA filtering with metabarcoding and undertook one of the most extensive lake sampling approaches to date, collecting 2 L samples every 1 km along the littoral zone, with further samples at each of these points from both 5 m and 20 m depth profiles into the limnetic zone. They found that eDNA was heterogeneously distributed and more species were detected from shoreline

samples. This type of intensive sampling strategy, combined with carefully implemented metabarcoding is likely to yield the highest probability of detection of biodiversity within an entire water body.

4.2.3 How is macrobial eDNA distributed within an aquatic environment?

Environmental DNA detectability is likely dependent on the interplay between DNA release and DNA degradation (Dejean *et al*. 2011; Thomsen *et al*. 2012a), which is affected by a suite of variables discussed in the Introduction (page 14). In summary, eDNA release rate is likely to depend upon organism size (Klymus *et al*. 2015; Lacoursière-Roussel, 2016b), and/or biological activity (Bylemans *et al*. 2016; Dunn *et al*. 2017), season (Goldberg *et al*. 2011; Vervoort *et al*. 2012; de Souza *et al*. 2016; Buxton *et al*. 2017b; Sigsgaard, *et al*. 2017; Stoeckle *et al*. 2017; Uchii *et al*. 2017), organism species density (Pilliod *et al*. 2013; Pilliod *et al*. 2014), DNA dispersal rates (Deiner *et al*. 2014; Taylor *et al*. 2015; Jane *et al*. 2015) and DNA or cell sloughing/shedding rate (Lacoursière-Roussel, 2016b; Sassoubre *et al*. 2016). Degradation rate of eDNA is likely to increase when environmental conditions have higher temperatures (Pilliod *et al*. 2014; Strickler *et al*. 2015; Eichmiller *et al*. 2016a; Lacoursière-Roussel; 2016b, Lance *et al*. 2017; Tsuji *et al*. 2017a), lower pH values (Seymour *et al*. 2018) increased exposure to ultraviolet light (Pilliod *et al*. 2014; Strickler *et al*. 2015), and increased bacterial and/or fungal action (Matsui *et al*. 2001; Dejean *et al*. 2011; Lance *et al*. 2017).

The distribution of eDNA is of particular importance for the development of effective monitoring methods (Darling and Mahon, 2011). It has been proposed for some time that organismal distribution may influence eDNA concentration within a water body (Takahara *et al*. 2012) and some recent studies have explored this topic. Eichmiller *et al*. (2014), using qPCR, showed local correlation of Common Carp (*Cyprinus carpio*) eDNA concentrations to 'high-use' and 'low-use' areas of a lake, indicating patchy distribution and possibly rapid eDNA degradation. Yamamoto *et al*. (2016) demonstrated local variation of eDNA concentrations in a marine bay in Japan, sampling in triplicates over a grid of roughly 400 m equidistant points across a ~ 10 km bay. This study recorded qPCR copy number of Japanese Mackerel (*Trachurus japonicus*) and found that it correlated well with echo sounder results, exhibiting highly localised eDNA concentrations, such as increased signal around the location of a wholesale fish market. Similar results were observed later by the same team using the exact same system, but for a jellyfish species, the Japanese Sea Nettle (*Chrysaora*

pacifica) (Minamoto *et al*. 2017). This study also observed significantly higher concentrations from samples taken 1.5 m above the sea floor, indicating that eDNA was likely localised according to the jellyfish's habitat preference, which may have been dictated by the deeper water habitat preference of its prey choice. Even more localised still, Davidson *et al*. (2017) showed variation in qPCR amplification of the invasive Asian cyprinid fish, Topmouth Gudgeon (*Pseudorasbora parva*) in an angling pond, with sampling sites spaced just 100 m apart along the shoreline.

Similar studies have also been conducted in marine habitats. O'Donnell *et al*. (2016) used 16S metabarcoding of metazoa along marine transects following an increasing depth gradient, and found distinct eDNA communities distributed in a non-random fashion. Port *et al*. (2016), using metabarcoding, demonstrated differences among marine fish communities sometimes separated by less than 100 m, revealing a correlation between community structures and specific habitat types. Kelly *et al*. (2018) recently found that nearshore organismal communities of benthic and planktonic taxa are largely consistent across tides, restricted to the site and water mass sampled, but as physiochemical water mass characteristics changed, the community composition of a broad range of organisms shifted in turn.

4.2.4 Challenges of sampling eDNA in the tropics

Tropical freshwater ecosystems have unique climatic challenges with regards to eDNA sampling. Sediment load and algal pollution creates higher than usual turbidity, caused by increased run off as a result of deforestation and conversion of natural landscapes to agricultural land (Asian Development Bank, 2016), and as the tropics are near the equator, they of course experience some of the highest global temperatures and UV light intensity. Information generated from eDNA collected in tropical biomes is therefore likely to differ in its ecological implications to that generated from colder biomes such as those in tundra, boreal or temperate regions. As eDNA degrades more rapidly with increased environmental temperature (Pilliod *et al*. 2014; Strickler *et al*. 2015; Eichmiller *et al*. 2016; Lacoursière-Roussel, 2016b; Lance *et al*. 2017; Tsuji *et al*. 2017a), and increased UV light (Pilliod *et al*. 2014; Strickler *et al*. 2015), it is expected that eDNA signals from tropical waters will represent a more contemporary 'snapshot' of native biodiversity,. There have been solutions suggested to deal with the specific challenges related to tropical eDNA sampling, including

the use of broad pore size filters (20 μm) (Robson *et al*. 2016), or storage buffers such as RNA Later (Ishige *et al*. 2017).

There have been a number of studies employing metabarcoding techniques in the tropics from both biological and environmental sources. Vietnamese forest mammals were detected from leech blood (Schnell *et al*. 2012), nematode diversity from Costa Rican rainforest microhabitats (Porazinkska *et al*. 2010), plant diversity from rainforest soil in French Guiana (Yoccoz *et al*. 2012), planktonic microbiota from Caribbean marine water (Rusch *et al*. 2007), and insects from Malaise traps in Malaysia (Ji *et al*. 2013). However, there have been few aquatic eDNA studies targeting tropical macrobial life. Piaggio *et al*. (2014) detected Burmese Python eDNA from waters in South Florida; Robson *et al*. (2016) detected the invasive Mozambique tilapia (*Oreochromis mossambicus*) in Northern Australia; Ishige *et al*. (2017) detected several endangered forest mammals from water surrounding salt licks in Sabah, Borneo, and Bakker *et al*. (2017) detected shark eDNA from Caribbean marine waters. Kapoor *et al*. (2017) also conducted basic population level analysis, assessing human haplotype variation from eDNA from watersheds in Puerto Rico. A very recent study used eDNA metabarcoding to monitor the fish community of a tropical lake in Mexico (Valdez-Moreno, 2018), but to our knowledge, this is the first ever to use aquatic environmental DNA to study a lake from 'mega-diverse' Southeast Asia, in particular, Indonesia.

4.2.5 Tropical lake ecology

Lacustrine habitats are generally self-contained, with specific habitat niches and community interactions. The cycle of water, nutrients, gas and light causes lake metabolism to fluctuate between anabolic photosynthesis and catabolic aerobic respiration (Likens, 2010). Food webs (see Figure 4.1 below) include both the benthic/littoral food chain, grazer food chain and microbial food chain (Likens, 2010).

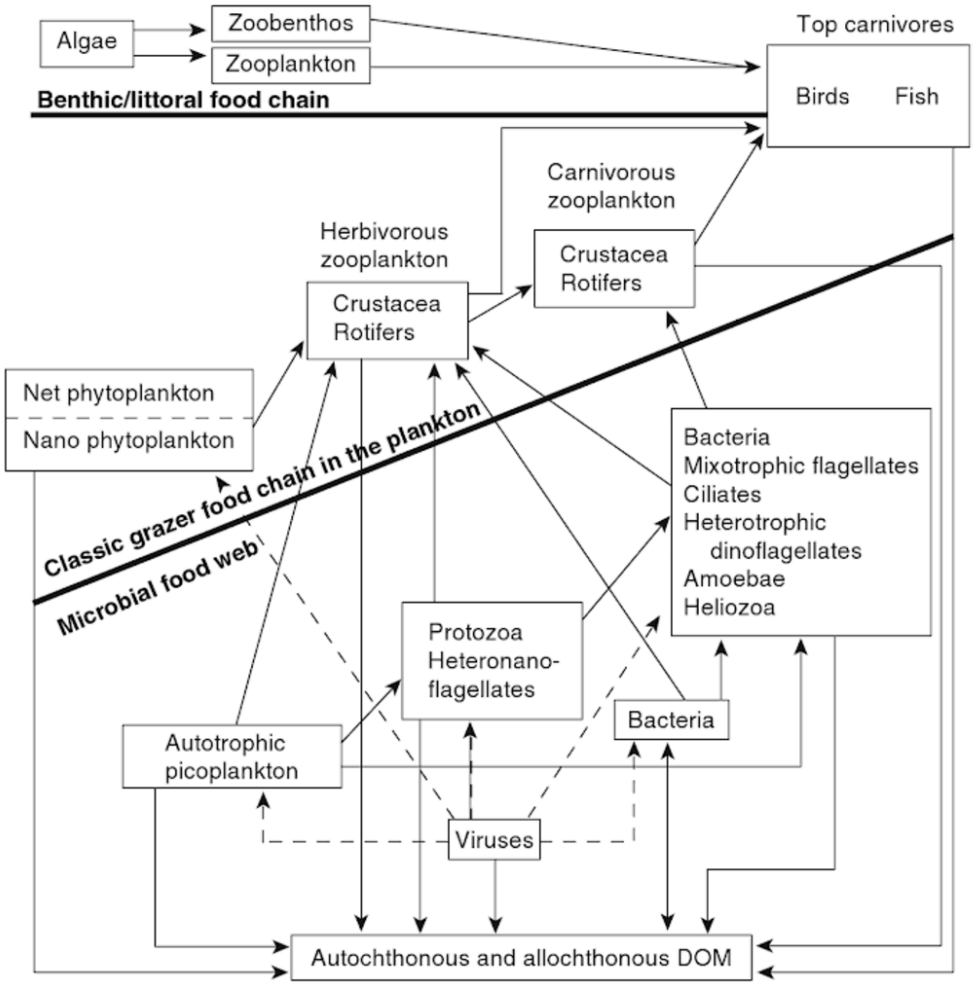


Figure 4.1. Diagrammatic view of a lacustrine food web. (from Weisse and Stockner, 1993 as modified by Kalff, 2002, taken from Likens, 2010).

Lakes with strong control by top carnivores are less responsive to nutrient input and subsequent problems of eutrophication, as larger fish control smaller fish, allowing large grazing zooplankton to thrive and subsequently control phytoplankton, resulting in a reduced response to nutrients such as phosphates and nitrates (Likens, 2010). Phosphorous and nitrogen are two elements most likely to be critically depleted by aquatic autotrophs, and so are thus commonly viewed as ecosystem regulators, with high amounts resulting in eutrophic lakes, and low amounts resulting in oligotrophic lakes (Likens, 2010). Consequently, human impacts through phosphorous loading via waste disposal, agriculture and soil disturbance has a negative effect on the trophic state, trophic web, and biodiversity of lakes. Varying physiological and abiotic factors cause the formation of lake zones (see Glossary and Figure 4.2 below).

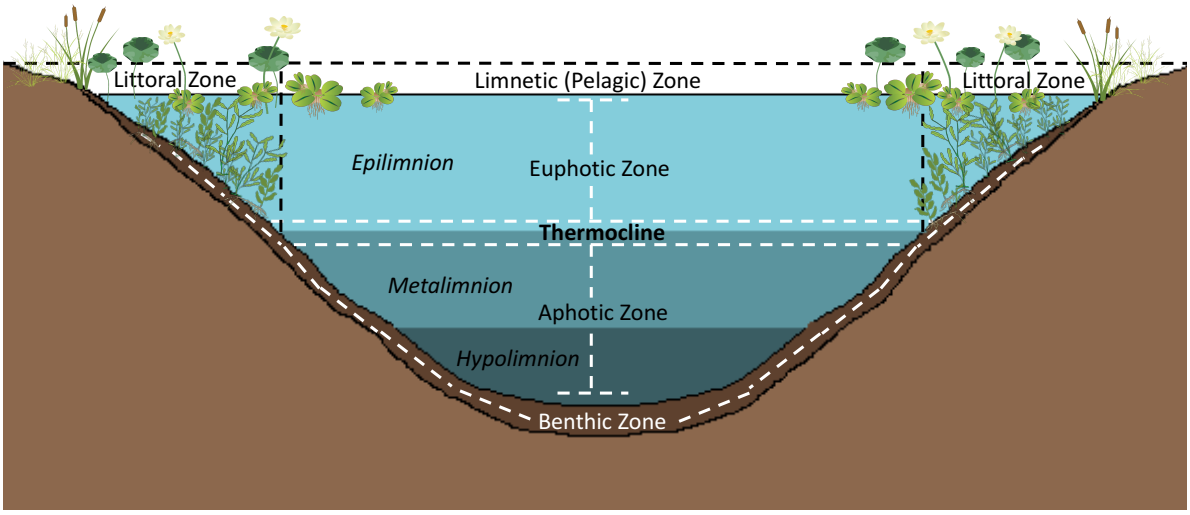


Figure 4.2 Diagram of lake zones. Shown here is the light-filled, plant-rich littoral zone and the open water, pelagic, limnetic zone in the centre of the lake. The limnetic zone is composed of different light zones, the light-filled euphotic zone within the epilimnion, and the dark aphotic zone across the metalimnion and hypolimnion. The euphotic zone and aphotic zone are separated by a thermocline at which point temperatures drop, and light decreases.

Phytoplankton and macrophytes can survive in the light-rich euphotic zone (usually in the epilimnion and sometimes the metalimnion), meaning that zooplankton communities differ between the littoral and limnetic zones as they feed on different prey. Fish communities may also differ between these zones due to the presence of structures around which to shelter from predation (Likens, 2010). If eDNA particles are heterogeneously distributed, as previous studies mentioned above suggest, then specific patterns of biodiversity and community structure should be observed from lake eDNA metabarcoding data according to the habitat type from which eDNA is collected.

4.2.6 Study site

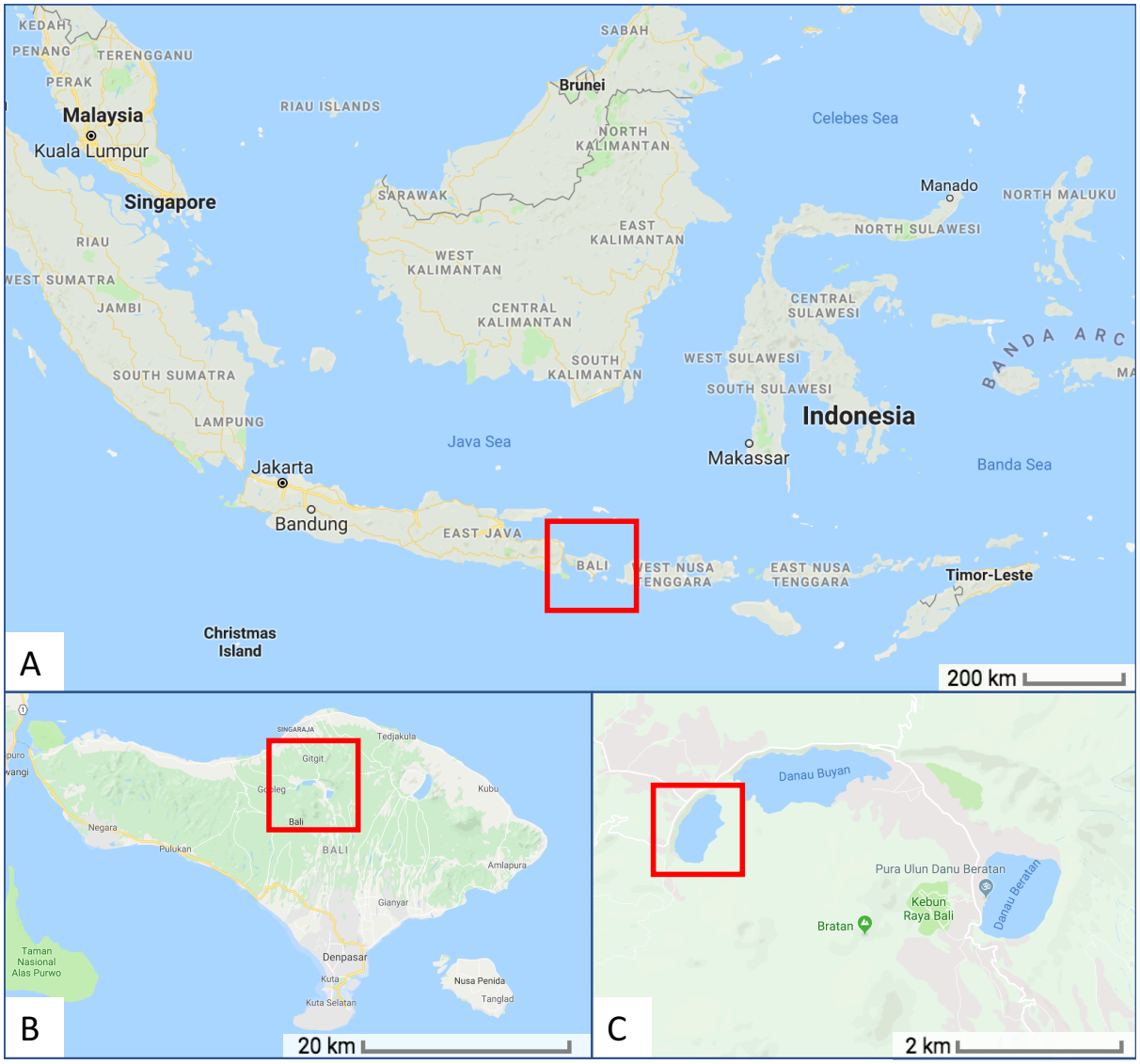


Figure 4.3 Location of Danau (Lake) Tamblingan. Locations are highlighted with red boxes. A: Bali, Indonesia. B the central mountain lakes of Bali. C: Danau Tamblingan.

Lake Tamblingan was selected out of the other Balinese lakes for its smaller surface area, minimising the sampling effort needed to cover the entire lake for such an intensive sampling approach as used for this experiment. Samples were collected on 04/07/2015 from Lake Tamblingan (Danau Tamblingan in the Indonesian language) Munduk Banjar, Buleleng Regency, on the island of Bali, Indonesia (S 8° 15' 26.96" E 115° 5' 46.852) (Figure 4.3). Bali is a tropical island, 8° south of the equator, with an average annual temperature of 27°C, average annual high of 30°C and average annual low of 25°C, and a defined rainy season November - March (Weatherbase, 2018a). Over the months of June and July 2015, Bali received 0 mm of rainfall (Weather Underground, 2018). Lake Tamblingan is a well-sheltered, meromictic, confined, land-locked, volcanic crater (caldera) lake at 1,214 m above

sea level (Lehmusluoto *et al*. 1997). As Lake Tamblingan is situated in this mountainous region, average temperatures are lower than those for Bali as a whole, with an annual average temperature of 23.1°C. It is the smallest of the Balinese confined lakes, with a surface area of around 1.9 km^2 and a maximum depth of 90 m (Lehmusluoto *et al*. 1997), although in this study, the deepest point detected using a remote depth detector was 36.5 m. It is the water reserve important for North Bali (Maghfiroh *et al*. 2016), and is situated amongst agricultural fields of rice, vegetables and coffee, the demand for which has resulted in some land areas being illegally cleared (Whitten *et al*. 1996). It is an important religious site, providing local income from visitors who require boat access by dugout canoe to nearby temples for religious activities and tourism (Lake Lubbers, 2018). The lake is permanently stratified into an oxygen depleted hypolimnion beginning at 29 m with a noticeably sharp secondary thermocline, and there is a gradient of electric conductivity between the surface and bottom (Lehmusluoto *et al*. 1997).

Lake Tamblingan has a Culture-Based-Fishery (CBF) involving stocking of hatchery-reared fish fingerlings into the waterbody. However, it is an oligotrophic lake, and due to its low productivity, interventions to attempt to introduce fish including Grass Carp (*Ctenopharyngodon idella*) have been made in the past without success (Whitten *et al*. 1996). News reports state that the lake is annually restocked with fingerling fish, claiming to maintain ecosystem health and support any fishing activities (Bali Travel News, 2016). In 2011 for example, the Fisheries and Marine Agency of the Buleleng Regency introduced roughly 200 'Ikan Bangeng' (The Milkfish, *Chanos chanos*), 10,000 'Ikan Karper' (Common Carp, *Cyprinus carpio*), 25,000 'Ikan Nila' (Nile Tilapia, *Oreochromis niloticus*) and 200,000 'Ikan Tawes' (Java Barb / Silver Barb, *Barbonymus gonionotus*) (Bulelengkab, 2013) to Lake Tamblingan and its neighbour, Lake Buyan. Lake Tamblingan is also one of the few lakes in the Buleleng Regency with a local fishery for catfish within the *Clarias* genus (Negara *et al.* 2015). However, the artificial stocking of fish for inland fisheries can have negative effects on coexisting fish biodiversity through demographic decline caused by waste and nutrient loading, predation on conservation-sensitive species, and fish escapes causing genetic contamination and introgressive hybridization of locally native fish (Thorpe *et al*. 2011; Anneville *et al*. 2015). Other threats to the lake and associated biodiversity include land use conversion, pollution, erosion and sedimentation and the introduction of alien species (Odada *et al*. 2005).

4.2.7 Aims and Objectives

This study aims to explore the spatial distribution of eDNA in a tropical lake, with implications for informing future sampling approaches. It is expected that an increase in the spatial intensity of sampling will in turn increase the amount of biodiversity associated information such as species richness and species diversity. However, it is unclear to what degree this could be observed in tropical lacustrine environments, and which areas of a lake should be prioritised. Here, aquatic eDNA samples were collected from Lake Tamblingan, at both the surface and different depths as described in the Methodology, and amplified using COI, 12S and 16S markers using metabarcoding and NGS to generate OTUs.

Aim 1: Assess the use of eDNA metabarcoding in recording species present in a tropical lake.
Objective 1a: Observe read counts per PCR for each marker, including the amount of reads assigned to Human DNA to explore amplification consistency and specificity.
Objective 1b: Observe the taxonomic assignments of the OTUs produced and compare with previously recorded species from the lake and local area, as well as the known distribution of these species or higher level taxa.

Aim 2: Assess whether species richness varies between different spatial points of the lake, i.e. between limnetic and littoral and shallow or deep lake zones.
Objective 2: Compare OTU richness (roughly equivalent to species richness) between a) Surface Lake Zones (limnetic *vs* littoral), and b) Sample Depths (Shallow *vs* Deep).
H_0 : There is no statistical difference between points sampled from different categories.
The null hypothesis is: H_0: $m_A = m_B$, where m_A and m_B indicate the group mean of OTU richness within each category.

Aim 3: Assess the spatial distribution of eDNA biodiversity information according to sample sites across the surface and depth gradient of the lake.
Objective 3: Compare community composition from OTUs and evidence for fine-scale community partitioning at different sites.
H_0 : There is no statistical difference in OTU communities between points sampled across the lake according to a) Lake Zone (limnetic *vs* littoral), or b) Sample Depth (depth at which the sample was collected).

4.3 Methodology

4.3.1 Collection of aquatic eDNA samples

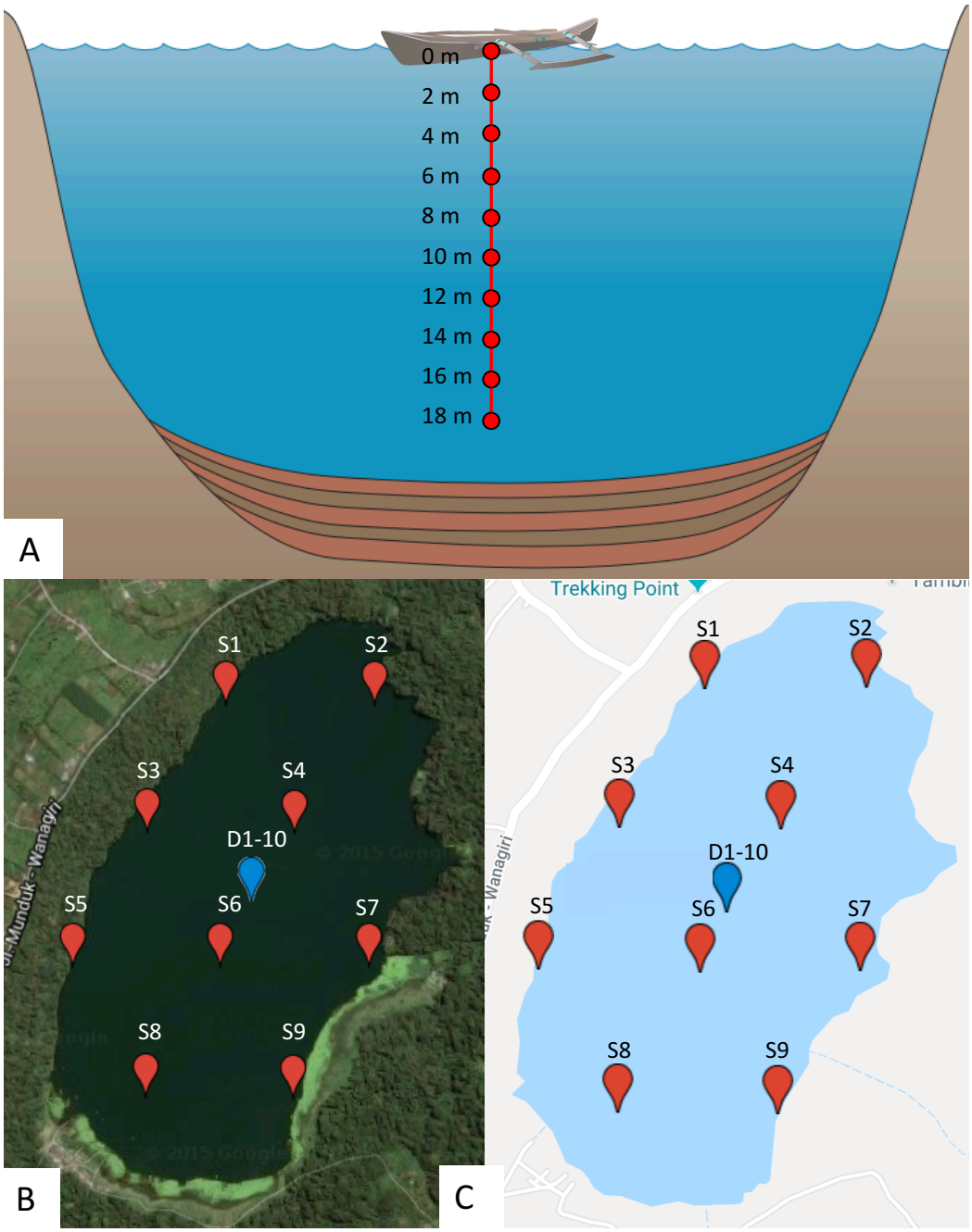


Figure 4.4 Lake Tamblingan sampling site. A: schematic diagram of depth sampling approach, with samples taken at 2 m intervals down to 18 m depth from a dugout canoe in the centre of the lake. B: Google Earth image showing surrounding forest, agricultural land, small settlement, and aquatic vegetation are visible. B&C: show surface sampling sites (red), and the point at which depth sampling was completed (blue). C: Google Maps image showing nearby roads and the location of a stream, indicated by the blue dotted line.

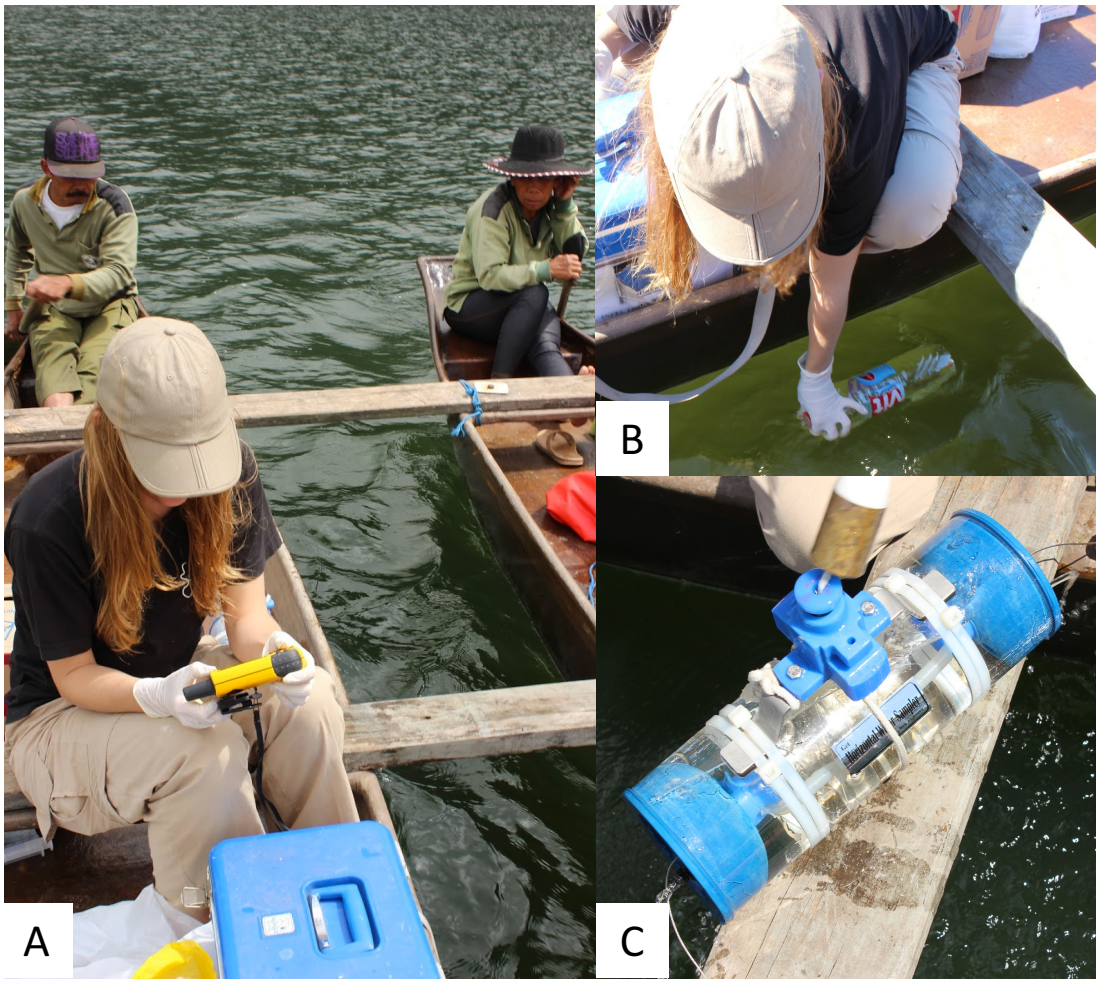


Figure 4.5. Lake Tamblingan sampling strategy. A: a GPS was used to access predetermined lake surface points via local dugout canoe. B: unused sterilised drinking water bottles were used to collect water from just below the surface. C: the Van Dorn horizontal water sampler employed for depth sampling.

Sampling points were decided upon in advance by assessing the lake size and layout using Google Earth, and were measured using Google Maps in the 'My Maps' application with the ruler tool which measures distance and areas (see Table 4.4.5 for GPS locations and environmental measurements). Points were selected to be 500 m apart, evenly spaced across the lake, and then entered into a GPS, accessed via local dugout canoe (Figure 4.5 A). Water was collected whilst wearing single-use nitrile gloves, in new 1.5 L plastic water bottles (Figure 4.5 B), (from which water was poured out and the outside of the bottles cleaned with 20% bleach and rinsed with ethanol) that had not been previously opened or stored in a molecular laboratory. Once at the exact sample point, the cap was removed, and the bottle dipped just below the surface until full, then the cap replaced (Figure 4.5 B). At each surface point, temperature, pH, dissolved oxygen and bottom depth were measured using digital

sensors, and turbidity was measured using a secchi disk and measuring tape. Nitrate and phosphate was also measured from each point using aquarium kits (detailed in the Universal Methods section). At the centre of the lake, a surface sample was collected using the above approach, and then further samples collected at different depths using a Van Dorn horizontal water sampler (Van Dorn Horizontal Water Sampler 2.2 L, model APAL - VHA 1, previously cleaned with 20% bleach and rinsed with ethanol). This was deployed and triggered to sample at increasing depths every 2 m until 18 m, which was the maximum depth possible using this equipment (Figure 4.5 C). Samples were collected in increasing order of depth so as to minimise mixing within the water column before sample collection. Once a sample was pulled up to the surface, it was poured into the same type of 1.5 L plastic bottles used to collect the surface samples. All water samples were filtered immediately after sampling at the shore of the lake as filtration and storage of samples on site immediately after collection is thought to best preserve eDNA yield (Spens *et al*. 2016; Yamanaka *et al*. 2016) (see 'Chapter 3: Universal Methods', 'Sterivex Filter Water Sampling'). Each 1.5 L bottle was first inverted several times to ensure homogenisation of eDNA particles, and was then sub-sampled by drawing up 3 sets of 500 mL of water at a time. Using a syringe, water was drawn up from the bottle and then pushed through a filter unit, so that there were 3 x Sterivex filter unit replicates containing eDNA from 500 mL of water, totalling 1.5 L from each sample point. A field blank was also collected by taking a Sterivex capsule out of the packaging for the same amount of time as it took to filter a single sample, and storing it in the same way as the other capsules during sampling, transportation, and laboratory storage. No distilled water was filtered through the field blank so as to minimise external sources of contamination. In total, there were ten sampling points along the depth gradient (0m – 18m, Figure 4.4 A) and nine different sampling points across the surface of the lake (S1-S9, Figure 4.4 B and C), each with three Sterivex filter replicates per point, yielding $n = 57$. Samples were stored in a standard freezer at -4°C overnight whilst in the field, and the next day placed in a -20°C freezer at the Indonesian Biodiversity Research Centre laboratory, Denpasar, Bali.

4.3.2 Molecular and bioinformatic methods

DNA was extracted from samples at the Indonesian Biodiversity Research Centre. DNA extractions were performed in a room where no PCR had ever been done, and on the floor of

a building where only microbiological extractions and associated PCR had been conducted. Before sample processing, the windows, walls, floor, surfaces and equipment were thoroughly cleaned with 20% dilution of commercial bleach solution, and rinsed with newly opened ethanol diluted with bottled drinking water to 70%. All laboratory consumables and equipment (Eppendorf tubes, racks, pipettes) were newly delivered and brought to Indonesia (except the vortex, centrifuge, incubation oven and freezer, although all were thoroughly cleaned as described above). After elution, 20 µl of each extraction was stored at the Indonesian Biodiversity Research Centre to allow collaborators access to samples, and the remaining 60-80 µl of each extraction shipped to The Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Denmark, for further analysis. The extraction, amplification, library building and bioinformatic analysis for all samples followed the approach which is detailed in the 'Chapter 3: Universal Methods'. Reads were filtered via several bioinformatic steps described in the 'Universal Methods', as well as a final baseline filter addition of 0.6%, 0.5% and 3% of the highest read count per OTU for 12S, 16S and COI respectively. This was decided upon based on the removal of spurious content in negative controls.

4.3.3 Descriptive statistics and statistical analysis

Firstly, the read count per PCR replicate was assessed and compared between PCRs. Secondly, the number of reads generated assigned to Human (*Homo sapiens*) DNA compared to non-Human DNA was described using a bar chart, to highlight the challenge of Human contamination in the eDNA molecular pipeline. Data was then placed into categories according to the marker used: 1) Total (all markers combined), 2) COI, 3) 12S and 4) 16S. OTU richness was compared between lake zones at the surface and different lake depths according to these four marker categories. Either an Unpaired Two-Sample T-Test, or an Unpaired Two-Sample Wilcoxon Test was used in R to compare species richness between a) Lake Zone (from surface samples): 'limnetic' (lake depth of >15 m, n = 5) or 'littoral' (lake depth of <15 m, n =4) samples, and b) Sample Depth (from depth transect samples): 'shallow' (sample depth of 0 - 8 m, n = 5) or 'deep' (sample depth of 10 - 18 m, n = 5). The use of these tests was decided upon after performing a Shapiro-Wilk Normality Test. Homoscedasticity (equal variance of each category) was not observed in any category according to depth sample data, nor in 12S and 16S surface sample data. Only Total and COI marker categories showed homoscedasticity in surface data, and therefore comparisons

within these groups were performed using an Unpaired Two-Sample T-Test, and all others were performed using an Unpaired Two-Sample Wilcoxon Test.

Non-Metric Multidimensional Scaling (NMDS) plots were created using Vegan (Oksanen *et al*. 2013) and ggplot (Wickham, 2016) in R using the Total Marker data category. Firstly, a Similarity Profile Analysis (SIMPROF) test for community structure using the Bray Curtis dissimilarity method for calculating a distance matrix was applied to the NMDS to observe patterns within the community composition of samples from each point on Lake Tamblingan. This analysis creates statistical clusters spatially superimposed upon NMDS data points to highlight data points which are statistically more similar. A Permutational Analysis of Variance (PermANOVA) was also performed on the NMDS OTU distance tables using ADONIS from the vegan package in R with 999 permutations. Secondly, NMDS plots were created in combination with the Manhattan dissimilarity method for calculating a distance matrix (chosen by Vegan based on the dataframe of these OTUs). These NMDS plots were created using total marker information and normalised read counts to the minimum read count (COI = 47,000 12S = 9,000 reads, 16S = 4,000 reads) to assess patterns in the data relating to the categorical variable of different lake zones or different sample depths. Two variables were separately incorporated, consisting of Sample Depth (the categorical variable of depth at which a sample was collected) and Lake Zone. Lake zones were categorised into either 'limnetic' (lake depth of >15 m, n = 5) or 'littoral' (lake depth of <15 m, n =4), and sample depths categorised into either 'shallow' (<3m depth, n = 11) or 'deep' (3 – 18m depth, n = 8) using all surface and depth transect samples combined.

4.4 Results

After extraction, each of the three subsamples collected at each location were combined into one extract sample, which was then PCR amplified independently three times, to create a total sample set of n=57. The 57 eDNA extracts amplified with varying success depending on the marker used. For COI (Leray *et al*. 2013), 57/57 extracts showed strong bands on the gel. For 12S (Valentini *et al*. 2016), 11/57 showed strong bands, 33/57 medium strength bands, and 13 showed weak bands. For 16S (Taylor, 1996), 13/57 showed strong bands, 35/57 medium strength bands, and 9/57 weak bands. After initial bioinformatic filtering of the data (Chapter 3: Universal Methods) to yield only high-quality identifiable sequence read counts (including filtering for reads only found in 2/3 PCR replicates) average read count per sample for the COI data was 62,005, and the minimum to maximum range was 47,425 - 79,006, with 0 reads present in the negative control. Reads were normalised to the minimum read count of 47,000. For read counts per PCR amplification, see Tables 4.4.1., 4.4.2 and 4.4.3. Average read count per sample for the 12S data was 38,839, and the minimum to maximum range was 8,753 to 189,827, with 0 reads present in the negative controls. Reads were normalised to the roughly minimum read count of 9,000. Average read count per sample for the 16S data was 3,937, and the minimum to maximum range was 235 to 10,153, with 0 reads present in the negative controls. Reads were normalised to 4,000, roughly the median read count, as the minimum read count for 16S was so low that normalising to this number may have compromised the details of OTUs with low read counts by transforming them into decimal numbers less than 1. For the 16S data, there were two samples (S4 and D6) which before removal of known contaminants, contained 153,828 and 131,465 reads respectively assigned to *Homo sapiens*. For these two samples, the final read count after all quality filtering including the removal of contaminants was 0 (see Figure 4.11 taxonomy bar chart, in which S4 and D6 were removed), and so these samples were removed from further analysis. The remaining high-quality sequences assigned to samples from Lake Tamblingan, were collapsed and quality filtered into a total of 40, 12 and 5 OTUs, for COI, 12S and 16S respectively. The species observed through the 12S and 16S metabarcoding data, compared to previously recorded species from Lake Tamblingan from the literature are shown in Table 4.4.7. As the COI data did not generate reliable vertebrate OTUs, this data was not included in this table. After filtering and removal of sequences which were obvious contaminants (Appendix 6), negative controls were blank.

Table 4.1. Read counts per PCR replicate of all samples for COI data. Read counts are comprised of the reads remaining after filtering described in the Universal Methods section (minimum copy number 50), before creating OTUs and before final manual filtering.

Sample name	Read count PCR1	Read count PCR2	Read count PCR3
DTAMD1	24,543	15,063	24,426
DTAMD2	23,956	14,767	19,380
DTAMD3	13,324	12,650	30,141
DTAMD4	15,265	19,804	18,014
DTAMD5	26,904	12,376	22,720
DTAMD6	19,688	19,496	24,643
DTAMD7	20,617	18,077	22,075
DTAMD8	23,935	18,148	17,970
DTAMD9	18,390	16,205	20,796
DTAMD10	22,212	7,836	17,853
DTAMS1	30,289	21,917	22,752
DTAMS2	28,271	26,840	24,738
DTAMS3	26,110	22,361	24,710
DTAMS4	22,900	11,180	20,676
DTAMS5	19,059	21,587	16,850
DTAMS6	29,500	11,638	20,903
DTAMS7	20,830	14,730	20,035
DTAMS8	25,747	22,791	27,663
DTAMS9	22,416	15,748	36,400
DTAMFNEG	57	56	164

Table 4.2. Read counts per PCR replicate of all samples for 12S data. Read counts are comprised of the reads remaining after filtering described in the Universal Methods section, (minimum copy number 20) before creating OTUs and before final manual filtering.

Sample name	Read count PCR1	Read count PCR2	Read count PCR3
DTAMD1	35,115	72,647	54,425
DTAMD2	32,907	44,044	38,103
DTAMD3	54,515	91,542	70,805
DTAMD4	54,967	100,662	76,007
DTAMD5	4,904	32,829	41,743
DTAMD6	23,457	15,054	41,134
DTAMD7	44,781	68,174	86,714
DTAMD8	13,686	26,706	54,609
DTAMD9	38,635	128,835	137,631
DTAMD10	27,793	41,092	64,948
DTAMS1	50,084	43,112	77,401
DTAMS2	17,335	26,978	34,616
DTAMS3	42,361	32,530	110,913
DTAMS4	42,116	23,648	51,318
DTAMS5	174,600	121,647	151,546
DTAMS6	43,474	23,518	26,461
DTAMS7	43,626	60,509	29,126
DTAMS8	8,739	48,668	35,823
DTAMS9	70,781	93,033	60,184
DTAMFNEG	879	5,529	5,206

Table 4.3. Read counts per PCR replicate of all samples for 16S data. Read counts are comprised of the reads remaining after filtering described in the Universal Methods section, (minimum copy number 20) before creating OTUs, and before final manual filtering.

Sample name	Read count PCR1	Read count PCR2	Read count PCR3
DTAMD1	131,346	95,659	86,049
DTAMD2	76,540	51,385	53,686
DTAMD3	150,917	76,192	30,315
DTAMD4	85,485	56,700	71,008
DTAMD5	42,099	23,921	20,066
DTAMD6	55,511	29,044	46,910
DTAMD7	121,906	74,408	86,374
DTAMD8	66,780	81,594	70,137
DTAMD9	113,449	122,671	120,129
DTAMD10	31,206	80,234	64,410
DTAMS1	97,200	67,285	44,179
DTAMS2	32,888	18,879	43,127
DTAMS3	77,897	66,511	52,078
DTAMS4	67,787	53,669	32,372
DTAMS5	128,471	71,427	79,300
DTAMS6	14,624	34,042	1
DTAMS7	72,256	24,789	62,764
DTAMS8	51,078	54,378	46,209
DTAMS9	28,294	23,416	32,877
DTAMFNEG	8,893	1,790	1,113

Table 4.4 Basic descriptive statistics of read counts per sample per PCR replicate. Standard deviation, variance, sum, mean and standard error of read counts per sample according to PCR replicate for each marker after quality control and filtering. 12S reads filtered for presence a minimum of 2/3 PCR replicates and 20 copies. 16S reads filtered for presence a minimum of 2/3 PCR replicates and 20 copies. COI reads filtered for presence a minimum of 2/3 PCR replicates and 50 copies.

PCR Replicate	PCR1	PCR2	PCR3
12S			
Sample Standard Deviation, s	36,023	34,793	35,324
Variance (Sample Standard), s2	1,297,679,010	1,210,537,921	1,247,811,367
Population Standard Deviation σ	35,063	33,865	34,382
Variance (Population Standard), σ2	1,229,380,115	1,146,825,399	1,182,137,085
Sum	823,876	1,095,228	1,243,507
Mean (Average)	43,362	57,644	65,448
Standard Error of the Mean (SEx̄):	8,264	7,982	8,104
16S			
Sample Standard Deviation, s	39,262	27,959	27,570
Variance (Sample Standard), s2	1,541,475,176	781,686,375	760,084,152
Population Standard Deviation σ	38,214	27,213	26,834
Variance (Population Standard), σ2	1,460,344,904	740,544,987	720,079,723
Sum	1,445,734	1,106,204	1,041,991
Mean (Average)	76,091	58,221	54,842
Standard Error of the Mean (SEx̄):	9,007	6,414	6,325
COI			
Sample Standard Deviation, s	4,545	4,853	4,796
Variance (Sample Standard), s2	20,659,518	23,549,101	23,002,477
Population Standard Deviation σ	4,424	4,723	4,668
Variance (Population Standard), σ2	19,572,175	22,309,675	21,791,820
Sum	433,956	323,214	432,745
Mean (Average)	22,840	17,011	22,776
Standard Error of the Mean (SEx̄):	1,043	1,113	1,100

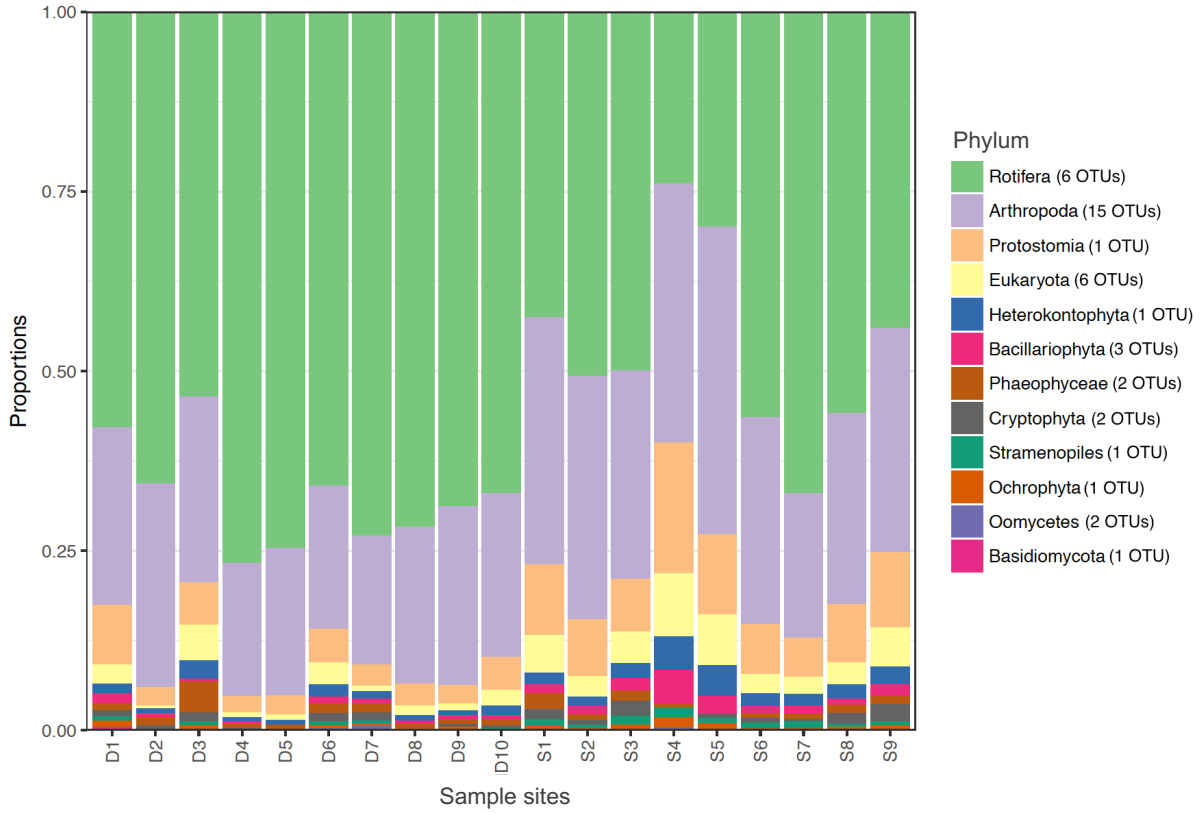


Figure 4.9 Bar chart of taxa per sample site from COI OTUs. The relative abundance of read counts per OTU from the COI metabarcoding data from Lake Tamblingan, assigned to phylum level, with the number of OTUs per phylum shown in the key.

The COI metabarcoding data using the Leray (2013) primers (described in the Universal Methods section) targeting a 313 bp region of the COI gene mostly amplified microfauna, meiofauna and microalgae, summarised in Figure 4.9 above. The number of OTUs per sample ranged from 17-32 with a mean of 26 (SD ±4.45) and a total of 40. OTUs 75, 138, 216, 57, 456 and 71 could only be assigned to the domain Eukaryota, and OTU 8 only to the unranked clade Protostomia, placed within bilateral animals. It was possible to assign the remaining OTUs to 7 taxonomic phyla, composed of Arthropoda, Basidiomycota, Cryptophyta, Heterokontophyta, Ochrophyta, Rotifera and Stramenopiles. These constituted 9 families, 3 genera, and only one species (OTU85, *Diaphanosoma excisum* freshwater ctenopod in the family Sididae). The highest number of unique OTUs were assigned to the phylum Arthropoda (13), followed by Rotifera (6) and Ochrophyta (5). However, the phyla that the majority of reads were assigned to were Rotifera (58%), Arthropoda (26%), and Prostisomia (6%).

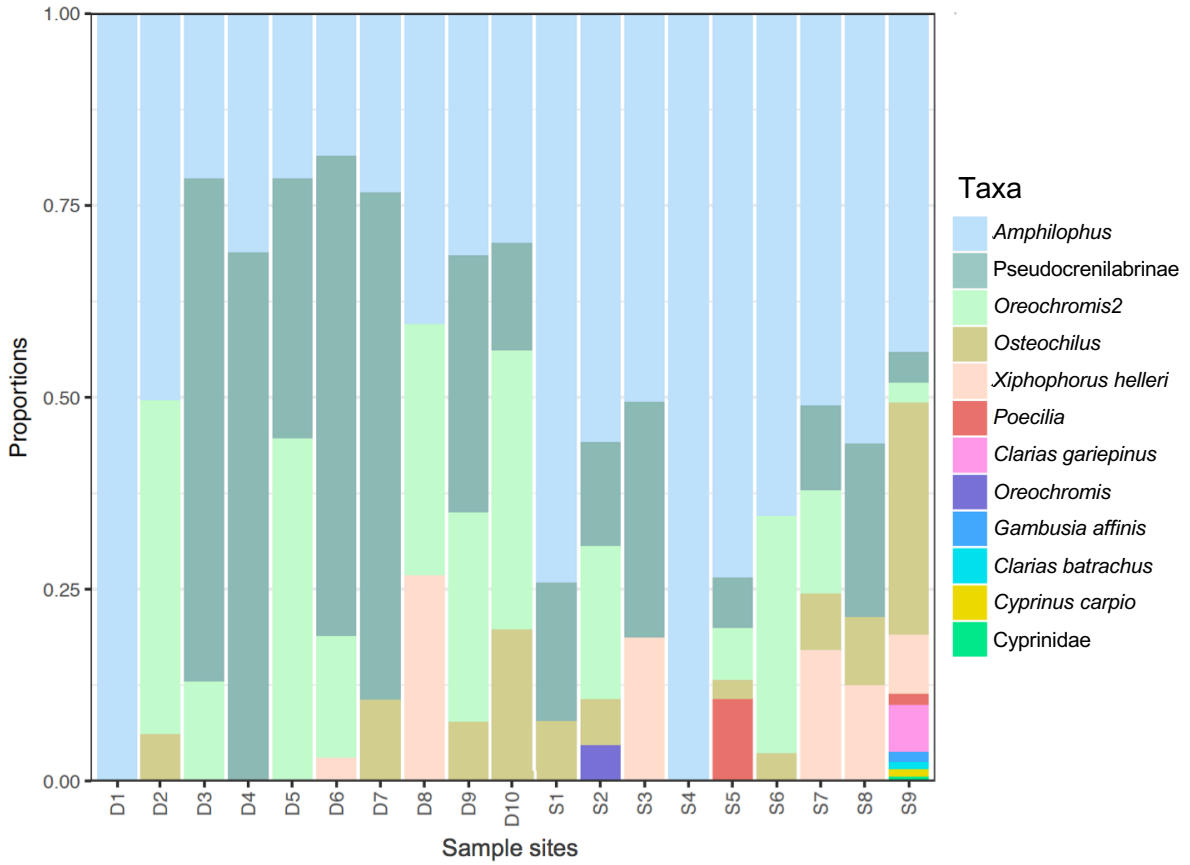


Figure 4.10 Bar chart of taxa per sample site from 12S OTUs. The relative abundance of read counts per OTU from the 12S metabarcoding data from Lake Tamblingan, assigned to either species or genus level.

For the 12S marker data, the number of OTUs per sample ranged from 1-11 with a mean of 4 (SD ±2.1) and total of 12, summarised in Figure 4.10 above. The 12S data created higher quality hits, with a Query Cover for all sequences of 100, and Identity ranging from 91 – 100. There was however, significant human amplification. After quality filtering, these 12 OTUs belonged to 6 taxonomic orders, 6 families, 9 genera, and five of the OTUs could be assigned to 5 species (*Cyprinus carpio*, *Gambusia affinis, Xiphophorus hellerii, Clarias gariepinus, Clarias batrachus*). Some of these OTUs were assigned to native fish (based on likely genera such as *Osteochilus*) known from this lake from a study in 1978 (Green *et al*. 1978) and some are additional species not described in this publication but known from the area (see Table 4.4.7) (no other literature than that mentioned in Table 4.4.7 was found concerning vertebrate species from this lake).

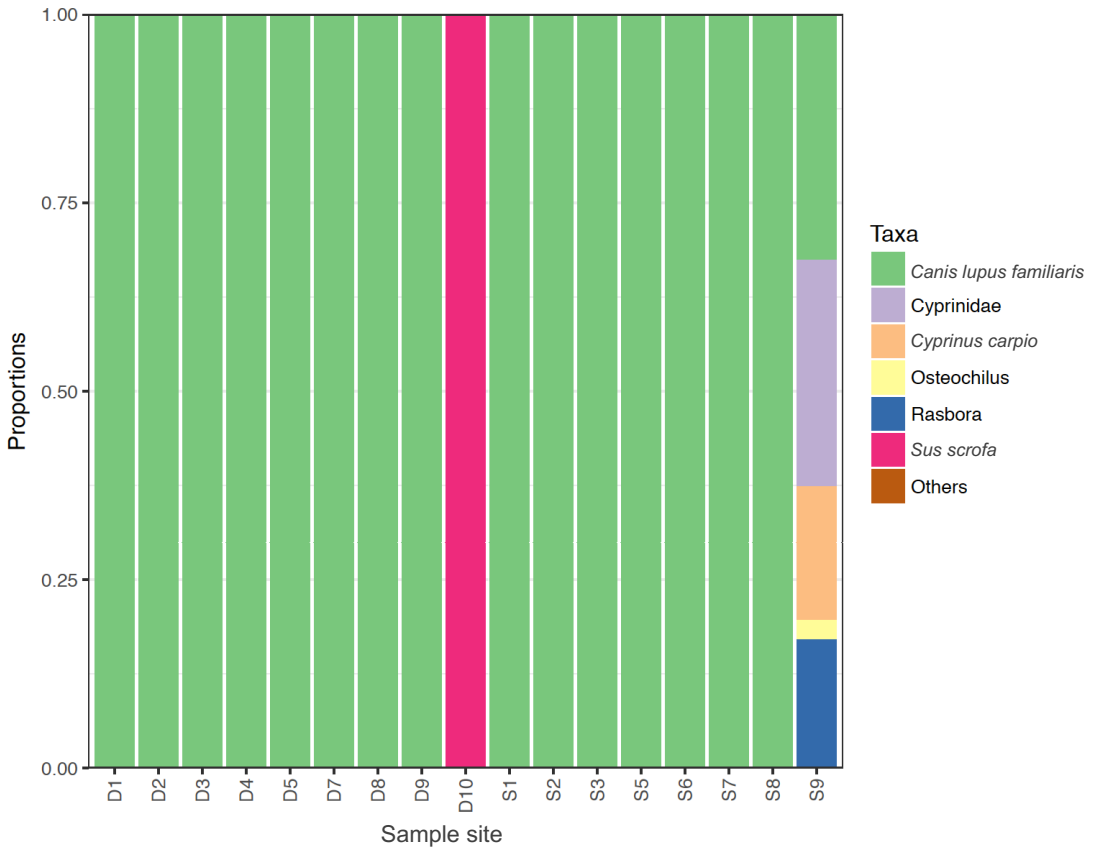


Figure 4.11 Bar chart of taxa per sample site from 16S OTUs. The relative abundance of read counts per OTU from the 16S metabarcoding data from Lake Tamblingan, assigned to either species or genus level.

The 16S marker data using the Taylor (1996) primers targeting a ~ 90 bp region of the 16S gene mostly amplified mammals, and also some fish, is summarised in Figure 4.11 above. This primer pair was the least successful in amplifying target eDNA. The number of OTUs per sample after all filtering ranged from 0-5 with a mean of 1 (SD ±1) and a total of 5. These were domestic dog (*Canis lupus familiaris*), cattle (*Bos taurus*), pig (*Sus scrofa*), Common Carp (*Cyprinus carpio*), and 3 fish sequences assigned to *Cyprinidae*, *Osteochilus*, and *Rasbora*. The 16S data generated higher quality hits than COI, all OTUs had a Query Cover 100 – and Identity from 94 - 100, although there was significant human amplification.

Table 4.5. Sample points and associated metadata

Sample	GPS Location	Temp	pH	DO	Turbidity (m)	Nitrate (mg/L)	Phosph (mg/L)	Bottom Depth (m)	Sample Depth (cm)	Lake Edge Habitat	Wider Habitat
S1	S 8 15.079, E 115 05.758	24.6	7.5	34.1	2.5	0-10	2	36.5	30	Rocky shore	Forest
S2	S 8 15.077, E 115 06.026	23.3	7.4	33	2.8		2	31.4	30	Rocky shore	Forest
S3	S 8 15.308, E 115 05.616	23.6	7.5	40	2.8		2	19.2	30	Rocky shore	Forest
S4	S 8 15.312, E 115 05.883	23.8	7.5	34.3	2.7		2	37.5	30	Rocky shore	Forest
S5	S 8 15.544, E 115 05.481	24.2	7.6	24	2.5		2	2.5	30	Rocky shore	Forest
S6	S 8 15.549, E 115 05.750	23.4	7.6	32.4	2.5		2	23	30	Rocky shore	Forest
S7	S 8 15.546, E 115 06.014	22.9	7.7	31.4	2.7		2	2.8	30	Aquatic plants	Grassy bank
S8	S 8 15.780, E 115 05.613	22.4	7.6	14.5	1.2		2	1.2	30	Aquatic plants	Grassy bank
S9	S 8 15.782, E 115 05.877	21.5	7.7	29	2.8		2	3.6	30	Aquatic plants	Grassy bank
D1	S 8 15.450, E 115 05.792	23.3	7.5	31	2.6		2	36.5	30	Open Water	Open Water
D2	S 8 15.450, E 115 05.792	NA	NA	NA	2.6	10	0.5-1	36.5	200	Open Water	Open Water
D3	S 8 15.450, E 115 05.792	NA	NA	NA	2.6		0.5-1	36.5	400	Open Water	Open Water
D4	S 8 15.450, E 115 05.792	NA	NA	NA	2.6		0.5-1	36.5	600	Open Water	Open Water
D5	S 8 15.450, E 115 05.792	NA	NA	NA	2.6		0.5-1	36.5	800	Open Water	Open Water
D6	S 8 15.450, E 115 05.792	NA	NA	NA	2.6		0.5-1	36.5	1000	Open Water	Open Water
D7	S 8 15.450, E 115 05.792	NA	NA	NA	2.6		0.5-1	36.5	1200	Open Water	Open Water
D8	S 8 15.450, E 115 05.792	NA	NA	NA	2.6		0.5-1	36.5	1400	Open Water	Open Water
D9	S 8 15.450, E 115 05.792	NA	NA	NA	2.6		0.5-1	36.5	1600	Open Water	Open Water
D10	S 8 15.450, E 115 05.792	NA	NA	NA	2.6		0.5-1	36.5	1800	Open Water	Open Water

Table 4.6. Vertebrate species recorded at Lake Tamblingan. Species found in the literature, and species or genus hits recorded by eDNA metabarcoding from the 12S and 16S region from this study (no vertebrates were detected using the COI marker). * = 20 – 99 reads; ** = 100-999; *** = 1,000 – 9,999 reads; **** = 10,000 – 99,999 reads; ***** = > 100,000 reads in total across all samples before normalisation but after filtering through the bioinformatic pipeline and custom 0.6% filtering step. Where there is a dash '-', no reads were observed.

Species / Genus	Reference	12S reads	Query Cover : Identity	12S samples	16S reads	Query Cover : Identity	16S samples
Amphilophus	-	*****	100 : 98	19/19	-	-	0/19
Anabas sp.	Green *et al.* (1978)	-	-	0/19	-	-	0/19
Barbonymus gonionotus	Bulelengkab (2013)	-	-	0/19	-	-	0/19
Barbodes microps	Green *et al.* (1978)	-	-	0/19	-	-	0/19
Canis lupus	-	-	-	0/19	*****		18/19
Channa striata	Green *et al.* (1978)	-	-	0/19	-	-	0/19
Chanos chanos	Bulelengkab (2013)	-	-	-	-	-	0/19
Clarias batrachus	Green *et al.* (1978)	***	100 : 100	1/19	-	-	0/19
Clarias gariepinus	Negara et *al.* (2015)	****	100 : 100	1/19	-	-	0/19
Cyprinus carpio	Green *et al.* (1978); Bulelengkab (2013)	***	100 : 100	1/19	***	100 : 100	1/19
Gambusia affinis	-	***	100 : 100	1/19	-	-	1/19
Monopterus albus	Green *et al.* (1978)	-	-	0/19	-	-	0/19
Oreochromis	Green *et al.* (1978); Bulelengkab (2013)	****	100 : 100	17/19	-	-	0/19
Osteochilus	Green *et al.* (1978)	****	100 : 100	11/19	**	100 : 98	1/19
Poecilia	Green *et al.* (1978)	****	100 : 95	2/19	-	-	0/19
Rasbora sp.	Green *et al.* (1978)	-	-	0/19	***	100 : 99	1/19
Sus scrofa		-	-	0/19	***	100 : 100	1/19
Xiphophorus hellerii	-	****	100 : 100	6/19	-	-	0/19
Xiphophorus maculatus	Green *et al.* (1978)	-	-	0/19	-	-	0/19

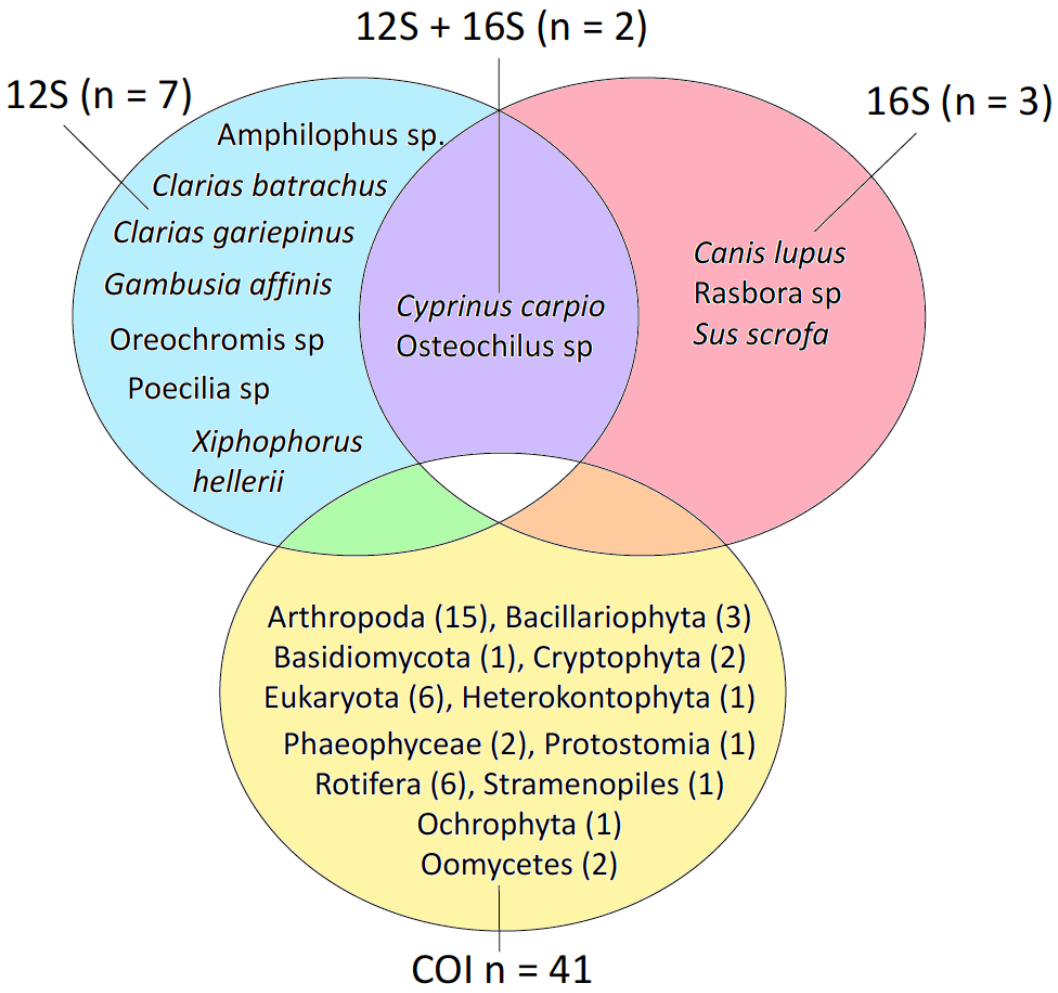


Figure 4.12. Venn diagram of species identified per marker. This Venn diagram shows species and higher level taxonomy identified by each primer, and where the same taxa were identified by multiple markers.

There was a small degree of overlap between the 16S and 12S primers which both amplified the Common Carp (*Cyprinus carpio*) and an OTU assigned to the *Osteochilus* genus. Apart from these two OTUs, all primers amplified a different range of taxa, with 12S mostly amplifying fish, COI mostly amplifying microfauna, meiofauna and microalgae, and 16S amplifying a small number of mammals and fish.

Table 4.7. Taxonomic information per sample. A) COI, B) 12S, C) 16S. S1 – S9 = surface samples (see Figure 4.4 A) and D1-D10 = depth samples (see Figure 4.4 B and C). Taxonomic assignments, sequence similarity and presence in a sample.

OTU		A: COI																					
	Phylum	**Lowest taxonomic rank**	**Taxonomic assignment**	**QC:ID**	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**
8	-	(clade)	Protostomia	71:79	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
75	-	Domain	Eukaryota1	91:81	■		■	■	■	■	■		■			■			■		■		■
138	-	Domain	Eukaryota2	97:99	■	■	■	■				■	■	■									
216	-	Domain	Eukaryota3	93:81	■		■	■	■	■				■									
57	-	Domain	Eukaryota4	94:81	■	■	■	■	■	■	■	■	■	■		■			■	■	■	■	■
456	-	Domain	Eukaryota5	98:81			■									■							
71	-	Domain	Eukaryota6	58:75	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	■	■	■
12	Arthropoda	Family	Cyclopidae	100:86		■				■		■	■		■	■	■	■	■	■	■	■	■
42	Arthropoda	Family	Cyclopidae	100:82	■	■	■		■	■	■			■	■								
85	Arthropoda	Species	*Diaphanosoma excisum*	99:99											■	■	■	■	■	■	■	■	
72	Arthropoda	Genus	Macrothrix1	91:93	■	■	■	■		■	■	■	■	■		■	■	■	■	■	■	■	■
4	Arthropoda	Genus	Macrothrix2	100:99	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
5	Arthropoda	Genus	Macrothrix3	99:92	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
88	Arthropoda	Genus	Macrothrix4	99:96	■	■						■	■										
149	Arthropoda	Genus	Macrothrix5	99:91	■	■	■	■					■	■									
516	Arthropoda	Genus	Macrothrix6	100:96									■										
363	Arthropoda	Genus	Moina	100:99		■										■							
522	Arthropoda	Order	Lepidoptera	83:78			■																
387	Arthropoda	Order	Lepidoptera	90:77			■					■											
18	Arthropoda	Order	Opiliones	67:74	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
191	Arthropoda	Superorder	Holometabola (Endopterygota)	89:80	■		■	■				■											
460	Basidiomycota	Genus	Rhodotorula	99:92										■									
129	Cryptophyta	Family	Cryptomonadaceae	95:81	■		■		■	■	■	■	■	■	■				■	■			
89	Cryptophyta	Family	Cryptomonadaceae	84:88	■	■	■			■		■	■	■	■	■	■	■	■	■		■	■
415	Heterokontophyta	Class	Oomycetes	97:90															■	■			

16	Heterokontophyta	Class	Oomycota	97:84	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
424	Heterokontophyta	Order	Peronosporales	89:73				■															
44	Heterokontophyta	Order	Thalassiosirales	99:89	■	■	■	■	■	■	■	■	■	■	■		■		■	■	■	■	■
102	Ochrophyta	Family	Bacillariaceae	97:89				■															
596	Ochrophyta	Family	Chordariaceae	93:82	■																		
47	Ochrophyta	Family	Dictyotaceae	72:82	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■
91	Ochrophyta	Order	Desmarestiales	95:73	■	■	■	■	■	■	■	■	■	■		■			■	■	■	■	
107	Ochrophyta	Order	Ectocarpales	83:75	■	■	■	■	■	■	■	■	■	■		■			■				
113	Rotifera	Class	Polyarthra	100:99					■														
170	Rotifera	Family	Floscularíidae	75:81		■									■		■			■	■	■	■
1	Rotifera	Order	Ploima1	98:84	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
120	Rotifera	Order	Ploima2	99:84	■																		
10	Rotifera	Order	Ploima3	95:83								■											
82	Rotifera	Order	Ploima4	98:86	■	■	■	■		■	■	■		■	■	■	■	■	■	■	■	■	■
90	Stramenopiles	Infrakingdom	Heterokonts	95:84	■	■	■	■	■	■	■	■	■	■		■			■	■			■
			OTU Richess	26	23	26	22	17	21	18	23	23	22	16	19	15	13	21	20	17	17	17	

Table 4.7. (continued) Taxonomic information per sample

Family	Lowest taxonomic rank	Taxonomic assignment	QC:ID	S1	S2	S3	S4	S5	S6	S7	S8	S9	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
B:12S																						
Cichlidae	Genus	Amphilophus	100:98	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Cichlidae	Genus	Oreochromis1	100:100		■																	
Cichlidae	Genus	Oreochromis2	100:100		■			■	■	■		■		■	■		■	■		■	■	■
Cichlidae	Subfamily	Pseudocrenilabrinae	100:100	■	■	■		■		■	■	■			■	■	■	■	■		■	■
Clariidae	Species	*Clarias batrachus*	100:100									■										
Clariidae	Species	*Clarias gariepinus*	100:100									■										
Cyprinidae	Family	Cyprinidae	100:91									■										
Cyprinidae	Species	*Cyprinus carpio*	100:100									■										
Cyprinidae	Genus	Osteochilus	100:100	■	■			■	■	■	■	■		■					■		■	■
Poeciliidae	Species	*Gambusia affinis*	100:100									■										
Poeciliidae	Genus	Poecilia	100:95					■				■										
Poeciliidae	Species	*Xiphophorus hellerii*	100:100			■				■	■	■						■		■		
OTU Richness				3	5	3	1	5	3	5	4	11	1	3	3	2	3	4	3	3	4	4
C: 16S																						
Canidae	Species	*Canis lupus*	100:100	■	■	■		■	■	■	■	■	■	■	■	■	■		■	■	■	
Cyprinidae	Species	*Cyprinus carpio*	100:100									■										
Cichlidae	Genus	Osteochilus	100:98									■										
Cyprinidae	Genus	Rasbora	100:99									■										
Suidae	Species	*Sus scrofus*	100:100																			■
				1	1	1	0	1	1	1	1	4	1	1	1	1	1	0	1	1	1	1
Total OTU Richness				30	29	30	23	23	25	24	28	38	24	20	23	18	17	26	24	21	21	21

After filtering, the number of unique OTUs varied between sample sites and between markers used. The COI data had the highest OTU richness, (average 19.7 OTUs per site), followed by 12S (average 3.05 OTUs per site), and 16S (average 1.05 OTUs per site). One site had many more OTUs than others across all markers (S9), and one site had much less than other sites (S4). The COI marker had the highest OTU richness, and 16S the lowest (Table 4.4.9 below).

Table 4.8 Summary statistics of OTU richness according to each marker category of all samples.

Marker	Highest OTU richness	Lowest OTU richness	Average OTU richness	SD of OTU richness
COI	26	13	19.7	±3.64
12S	11	1	3.05	±1.31
16S	4	0	1.05	±0.77
Total	38	17	24.5	±4.86

Although COI generated the highest OTU richness, this marker produced the lowest quality BLAST hits (Table 4.4.10 below). The 12S and 16S markers both produced high quality BLAST hits ranging from 91 – 100% identity (Table 4.4.10 below).

Table 4.9 Summary of OTU BLAST results for query cover and identity for each marker.

Marker	Average Query Cover	Highest Query Cover	Lowest Query Cover	Average Identity	Highest Identity	Lowest Identity
COI	92	100	58	86	100	73
12S	100	100	100	99	100	91
16S	100	100	100	99	100	98

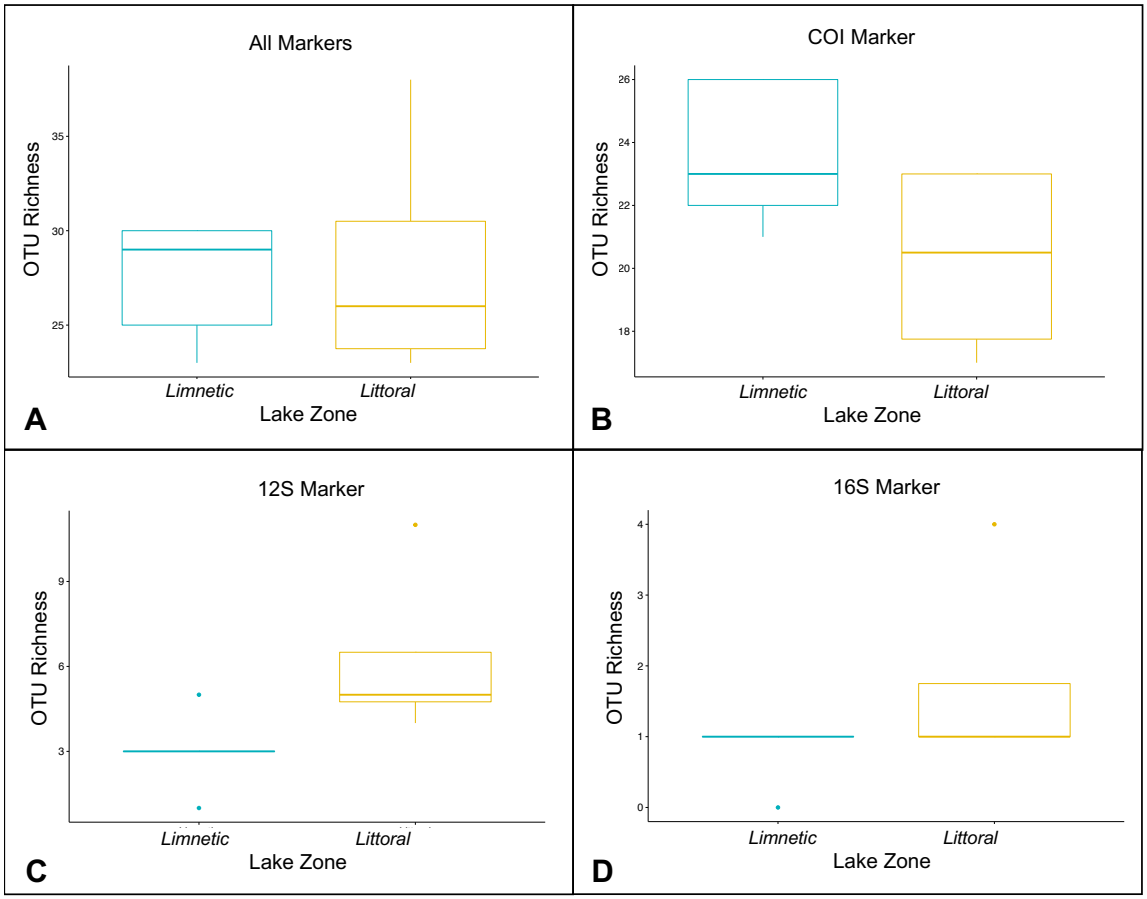


Figure 4.13. Boxplots of OTU richness by lake zone for all markers. OTU richness from surface samples categorised into either 'limnetic' (lake depth of >15 m, n = 5) and 'littoral' (lake depth of <15 m, n = 4) lake zones. A = all markers, B = COI marker, C = 12S marker, D = 16S marker.

Based on the Two-Samples T-Test, there was no significant difference between Littoral and Limnetic (Figure 4.13) samples for Total (p-value = 0.8109) or COI (p-value = 0.1094) data. Based on the Unpaired Two-Samples Wilcoxon Test, there was no statistically significant difference for 12S (p-value = 0.0572) or 16S (p-value = 0.240) data. There was a slightly higher average OTU richness in the limnetic zone than the littoral zone for COI, but a slightly higher OTU richness in the littoral zone for the 12S and 16S data (Table 4.4.11 below).

Table 4.10 OTU richness for each marker category according to lake zone.

Marker	Total		COI		12S		16S	
Lake Zone	Limnetic	Littoral	Limnetic	Littoral	Limnetic	Littoral	Limnetic	Littoral
Count	5	4	5	4	5	4	5	4
Mean	27.4	28.2	23.6	20.25	3	6.25	0.8	1.75
Standard Deviation	3.21	6.85	2.302173	3.201562	1.41	3.2	0.447	1.5

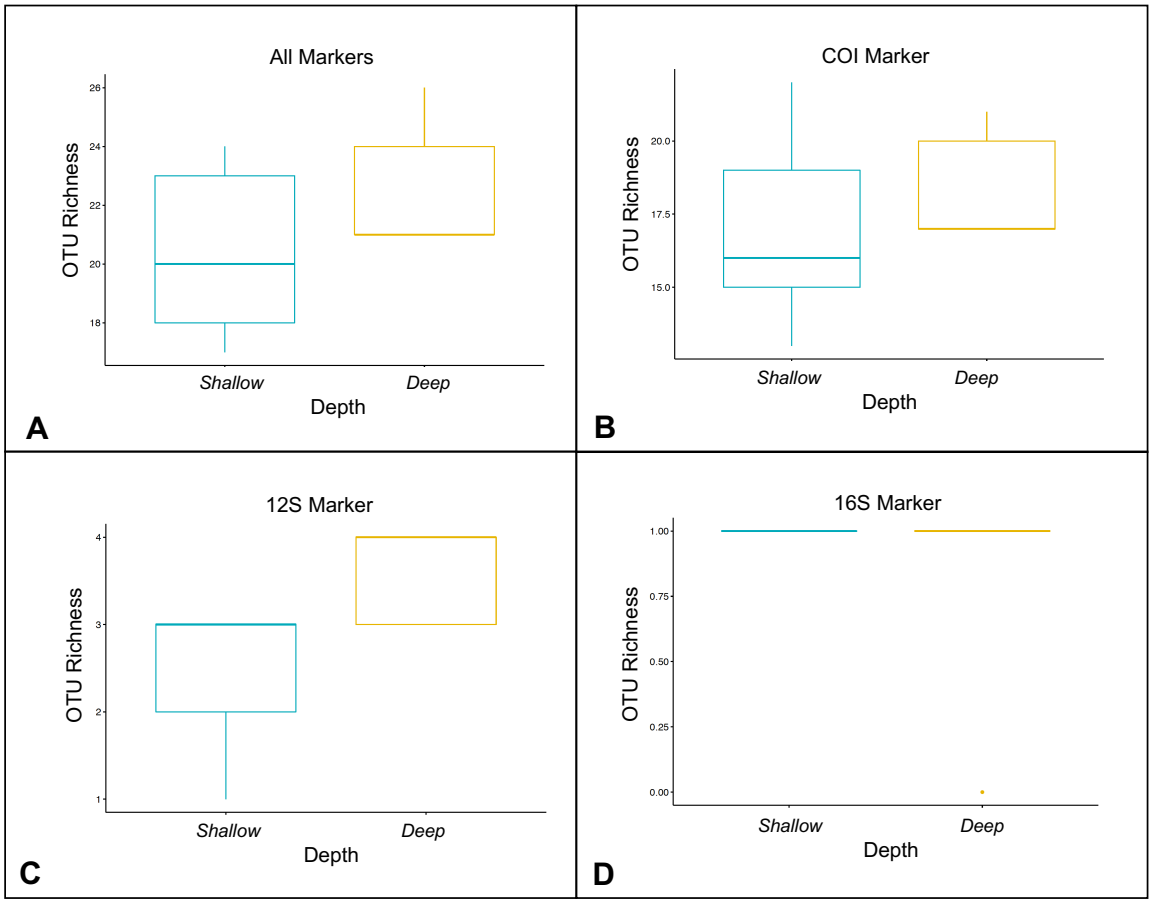


Figure 4.14. Boxplots of OTU richness by sample depth for all markers. Boxplots of OTU richness from samples collected at different depths, categorised according to depth zone. Samples were categorised into either 'shallow' (sample depth of 0 - 8 m, n = 5) or 'deep' (sample depth of 10 - 18 m, n = 4).

Based on the Unpaired Two-Samples Wilcoxon Test, there was no significant difference between Shallow and Deep (Figure 4.14) samples for Total (p-value = 0.243), COI (p-value = 0.397) or 16S data (p-value = 0.4237). There was however a statistically significant difference in the 12S Marker data between Shallow and Deep samples (p-value = 0.0419), in which there was a slightly higher OTU richness from Deep samples.

Table 4.11 OTU richness for each marker category according to lake depth.

Marker	Total		COI		12S		16S	
Depth	Shallow	Deep	Shallow	Deep	Shallow	Deep	Shallow	Deep
Count	5	5	5	5	5	5	5	5
Mean	20.4	22.6	17	18.4	2.4	3.6	1	0.8
Standard Deviation	3.05	2.3	4.54	1.95	0.894	0.548	0	0.447

There were seven significant clusters created in the NMDS plot (Figure 4.15 below), with a total stress value of 0.137 (an NMDS ordination with a stress value closer to 0.05 indicates a good fit, and closer to 0.3 indicates an arbitrary ordination, therefore 0.137 is a fair stress value).

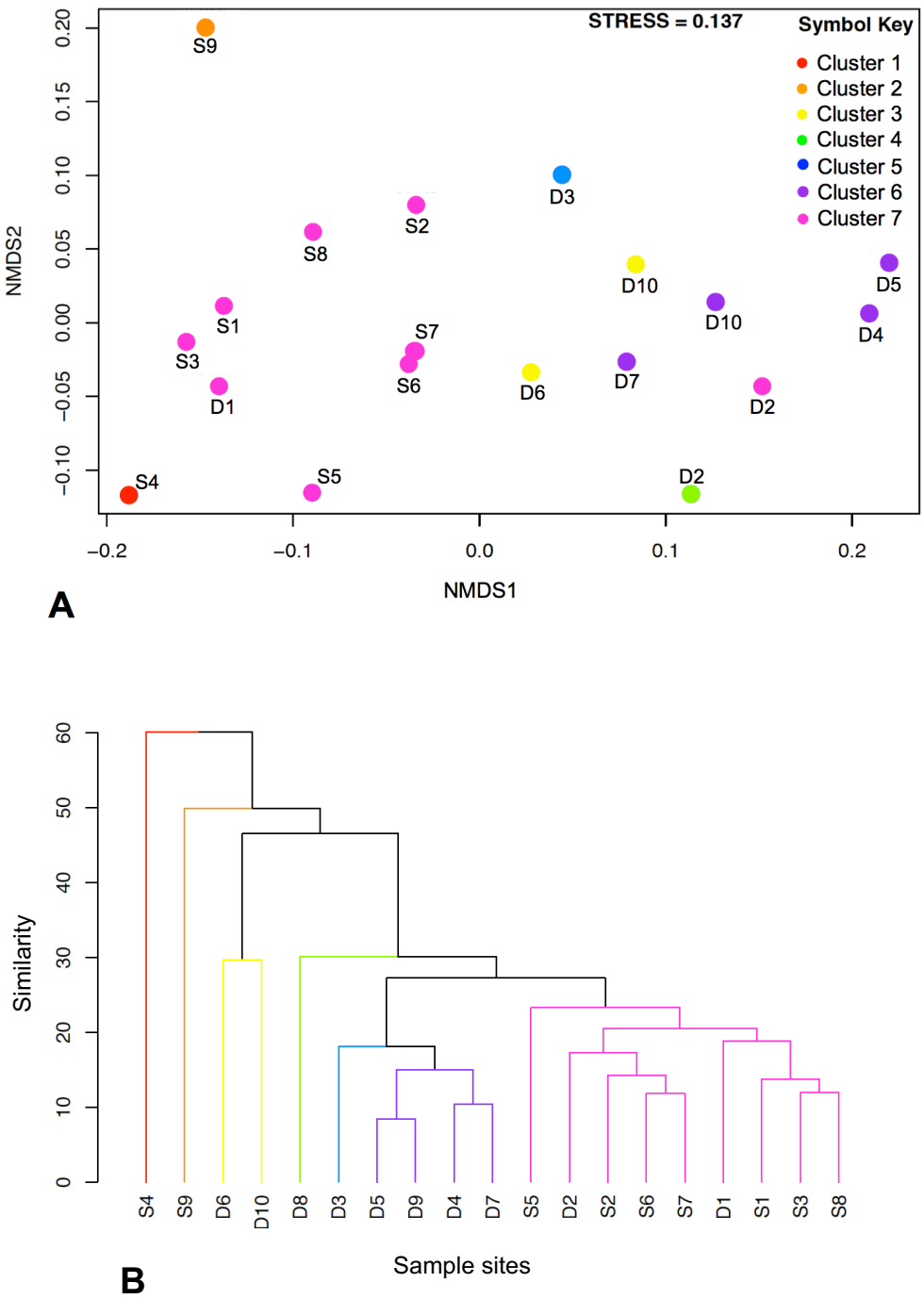


Figure 4.15 NMDS plot and SIMPROF dendrogram of all OTUs from samples. (labelled DTAM, meaning Danau Tamblingan, followed by the site name) and markers combined (Total) per sampling point of Lake Tamblingan using normalised read counts (top) and the SIMPROF dendrogram showing significant community clusters (bottom).

Figure 4.16 below shows a map of Lake Tamblingan with surface sample points, and a schematic diagram of the depth sampling points, both with colours corresponding to the SIMPROF analysis overlaid. This figure highlights that the surface, and shallow samples, exhibited statistically significant clustering, and that deeper samples clustered separately.

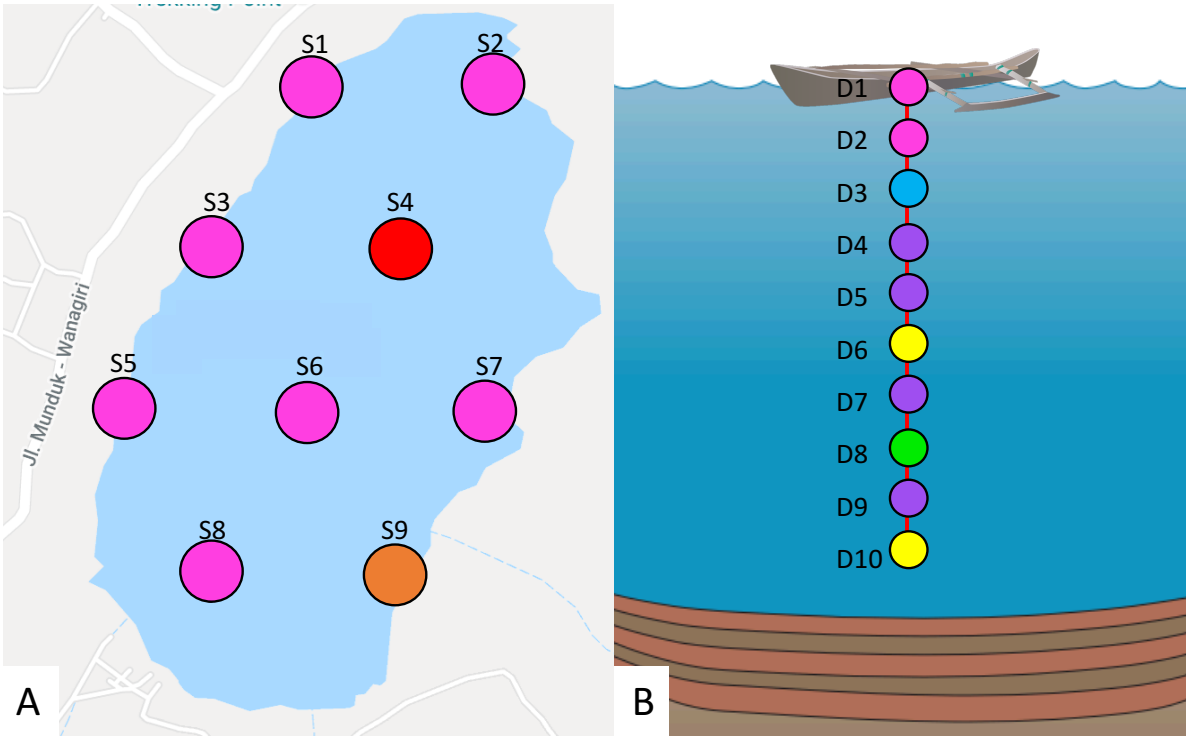


Figure 4.16. Lake Tamblingan sample points with NMDS SIMPROF clusters. Map of Lake Tamblingan and diagram of the different sampling points along the depth transect with the SIMPROF statistical clusters imposed upon each point.

The distance matrix derived from the Manhattan method showed no statistical impact of lake zones (Adonis PERMANOVA, $R^2 = 0.10$; $P = 0.089$), however there was a statistical impact of sample depth (Adonis PERMANOVA, $R^2 = 0.22$; $P = 0.004$).

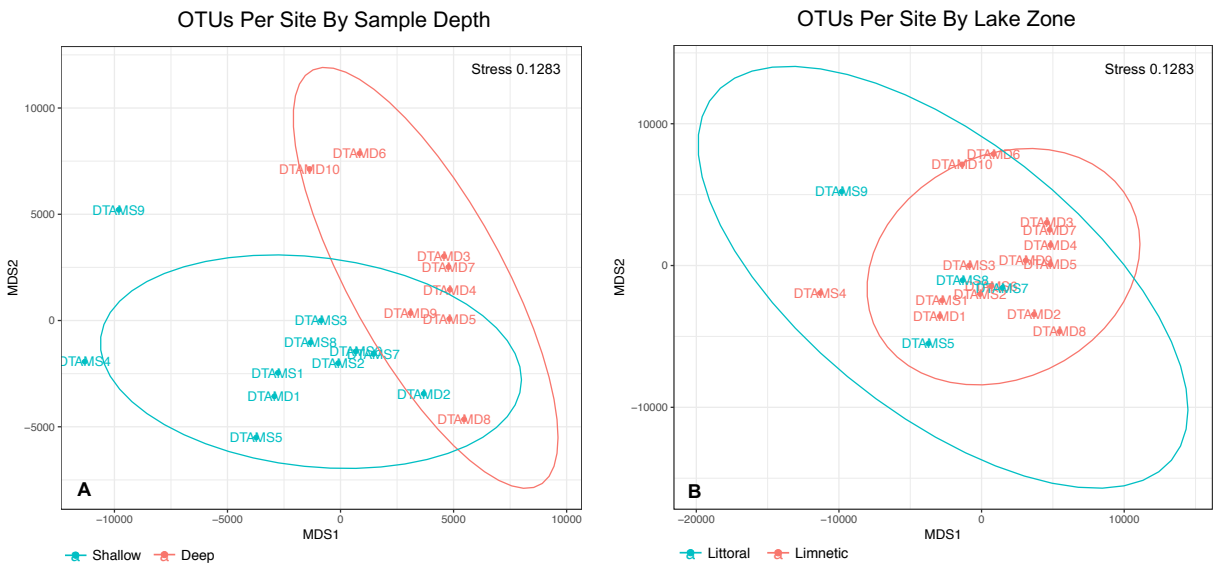


Figure 4.17. NMDS plots of community composition according to lake variables. A: Sample Depth and B: Lake Zone. Community composition is based on the normalised read counts per OTU of all markers combined (COI, 12S, 16S) with respect to A: all reads for all markers with respect to sample depth, B: all reads for all markers with respect to lake zones (limnetic or littoral). Ellipses signify the automatically generated clusters when dividing data according to A: Shallow vs Deep and B: Littoral vs Limnetic.

4.5 Discussion

4.5.1 Addressing Aim 1: Assess the use of eDNA metabarcoding in a tropical lake.

The first aim of this study was to assess the use of eDNA metabarcoding in recording species present in a tropical lake. The sampling of eDNA from Lake Tamblingan using Sterivex filters stored on ice had some advantages and some challenges. The Sterivex filters allow aquatic eDNA samples to be collected and filtered immediately using sterile syringes, meaning that there is no opportunity for cross contamination between samples, or from equipment, and that particles are captured from the water as quickly as possible. However, the fine pore size of the Sterivex filters mean that a large amount of human effort is necessary to force the desired volume of water across the filter membrane, which could only filter 500 ml per filter. Larger pore size filters may have been more suitable to this type of tropical lake to allow a greater volume to be filtered. In addition, storing the filters on ice, and in domestic style freezers overnight whilst the sampling trip was undertaken may have caused eDNA to degrade, or be more prone to amplify human contamination. When comparing the experimental samples from Lake Tamblingan with the positive control samples collected at the Anglesey Sea Zoo, there was a much higher proportion of human DNA in the Lake Tamblingan samples. The positive control samples were collected in containers, water stored in the freezer the same day, and then filtered as soon as the containers had defrosted (although this could also have been because of the higher density of non-human target eDNA and slower degradation in a temperate aquarium environment compared to a wild tropical environment where there was lots of human activity).

In terms of PCR amplification, variation was observed in the mean read count, and standard deviation between PCR replicates for each marker, shown in Table 4.4.4. The same samples were sequenced for each PCR replicate, using a different combination of PCR indexes and library indexes, and so the variation in read count may be due to this approach, or simply PCR stochasticity. This highlights the need to sequence PCR replicates separately to try to account for stochasticity within DNA amplification.

The presence of human DNA was high in lake samples amplified with the 12S and 16S markers, although considerably lower in positive control samples from the Anglesey Sea Zoo and Chester Zoo (the positive control samples were processed in the laboratory facilities at the Molecular Ecology and Fisheries Genetics Laboratory, Bangor University, details of which are described in Chapter 3: Universal Methods). This was unexpected, as the laboratory facilities at Copenhagen University employ a more stringent approach to laboratory cross contamination, including separate rooms for pre and post-PCR, unlike the

facilities at Bangor University. Another possibility is that the collection and filtering of the positive control samples by immediately freezing the water in containers after collection, and then immediately filtering after defrosting, may have better preserved the non-human target eDNA better than storing the eDNA samples from Sterivex filters on ice and in domestic freezers until arriving at the laboratory where filters could be stored at -20°C.

Seven out of the fourteen fish taxon recorded from Lake Tamblingan (Green *et al*. 1978; Negara *et al*. 2015; and Bulelengkab, 2013) were recovered using this eDNA metabarcoding approach. These included The Walking Catfish (*C. batrachus*), the African sharptooth catfish (*C. gariepinus*), the Common Carp (*Cyprinus carpio*), two OTUs from the *Oreochromis* genus, one OTU from the *Osteochilus* genus, one OTU from the *Poecilia* genus and one OTU from the *Rasbora* genus. In addition to these taxa which had previously been recorded in the literature, further OTUs were retrieved and from assigned to the *Amphilophus* genus, the species *Gambusia affini*, and the species *Xiphophorus hellerii*. There were few mammal species detected, but the OTUs assigned to mammal species that remained after filtering made biological sense (Domestic Dog - *Canis lupus* and Wild Boar - *Sus scrofus*). Over half of the previously recorded fish species were found, as well as other fish taxa (*Amphilophus*, *Gambusia affinis* and *Xiphophorus hellerii*) which made biological sense based on their distribution and description in the literature (Eidman, 1989; Siriwardena, 2010; Sentosa and Wijadi, 2012; Sentosa *et al*. 2013; Dahruddin *et al*. 2016), further discussed below. It is impossible to know whether the undetected species listed in Table 4.4.7 that were recorded by Green *et al*. (1978) and Bulelengkab (2013) (*Anabas* sp, *Barbonymus gonionotus*, *Barbodes microps*, *Chanos chanos, Monopterus albus, Xiphophorus maculatus*) were missing from the eDNA metabarcoding data because of a change in species composition since that study was conducted, or because the approach employed herein did not detect them (i.e. a false negative result).

Overall, although the eDNA metabarcoding of this aquatic lacustrine waterbody using this approach recovered species expected from this lake, this study would have benefited from a recent biodiversity surveys of this lake, using traditional methods for comparison. Due to the patchiness of the species distribution observed here, it is likely that the sampling and PCR approach (i.e. by doing PCRs on individual eDNA replicate extracts) would need to be increased to ensure accurate detection of total biodiversity.

4.5.2 Addressing Aim 2 - Compare OTU richness between different lake areas

The second aim was to compare OTU richness between different areas of Lake Tamblingan based on the surface samples from different lake zones (littoral *vs* limnetic) and different depths (shallow *vs* deep).

There was no significant difference in OTU richness between Littoral and Limnetic samples for any of the marker combinations. However there was a slightly higher OTU richness in the littoral zone than the limnetic zone for the 12S and 16S data, which aligns with the known ecology of lakes as discussed in the introduction, that fish and other animals use the cover of vegetation in the littoral zone and so spend more of their life cycle in this zone of the lake.

There was no significant difference in OTU richness between Shallow and Deep samples for the Total, COI or 16S data, but there was for the 12S data in which there was a slightly higher OTU richness from Deep samples. As the COI marker mostly amplified microfauna, meiofauna and microalgae which includes phytoplankton and zooplankton, it would be expected that there would be higher species richness in shallow waters (euphotic zone) than deeper waters, as phytoplankton and macrophytes (prey for zooplankton) survive best in the light-rich euphotic zone. However, as the depth sampler was only deployed to the maximum of 18m (due to the length of the rope used), only the euphotic epilimnion was sampled. This may indicate that if depth samples had been collected past the thermocline of 29m into the aphotic zone, a difference in community structure could have been observed. The slightly higher OTU richness for the 12S data from Deep samples may be due to less eDNA degradation in the darker, colder depths, or this pattern could have just occurred by chance due to a limited number of samples.

The null hypothesis, that there is no statistical significance between groups of OTU richness within each category can there for be accepted in terms of Lake Zones for all markers, and in terms of sample depth for all markers apart from 12S. A greater number of samples, with more reliable preservation and individually amplifying filter replicate extractions may have better illuminated these observed patterns.

4.5.3 Addressing Aim 3 – Assess evidence for fine-scale spatial community partitioning

The third aim was to assess the spatial distribution of eDNA and evidence for fine-scale community partitioning according to sample sites across the surface and depth gradient of the lake. The NMDS plot combined with the SIMPROF Analysis indicated that there were seven statistically separate community clusters (Figure 4.15 and 4.16). The samples which did not

fall into group clusters in the NMDS plot and SIMPROF dendrogram were S4 and S9 (red and orange in Figure 4.15 and 4.16). Sample S9 had the highest OTU richness for all markers, and sample S4 had the lowest for 12S and 16S, which may explain their separation. Site S9 was close to a stream, and so may have a unique community profile and higher species richness as the stream may host more fish species than the lake. The largest significant cluster is shown in pink, composed of 9 samples which were all either surface samples (S1, S2, S3, S5, S6, S7, S8, D1) or samples from near the surface (D2). The second largest cluster was composed of 4 samples, all from deeper depths (D4, D5, D7, D9) shown in purple, the third largest cluster was composed of 2 samples, both from deeper depths (D6 and D10) shown in yellow, and the final two clusters were only composed of one sample each - D8 (green) and and D3 (blue), which both clustered nearer to the deep samples.

If there was no difference in the community composition across spatial points, only one statistical cluster would have been found using the SIMPROF analaysis. The seven unique clusters shows that there was a difference in community composition based on the OTU table created from combining COI, 12S and 16S data. The clusters do make ecological sense, in that the largest cluster grouped the majority of surface samples together, whilst other clusters were formed from nearby depth samples at increasing depths.

The PERMANOVA test comparing shallow *vs* deep, and limnetic *vs* littoral communities from the combined reads assigned to OTUs of all markers (shown in NMDS plots in Figure 4.17) showed that there was a statistical impact of sample depth, but not of lake zones. It would be expected that communities from littoral and limnetic zones would differ due to the higher use of littoral habitats for structural shelter, however it may be that the sample collection approach or quality of eDNA preservation was not sufficient to observe significant differences.

4.5.4 Laboratory contamination

Low counts of reads removed in the custom % background filtering consisted of those assigned to the Kissing Gourami (*Helostoma temminkii*), the glass fish genus *Ambassis,* the Spanner Barb (*Barbodes lateristriga*) and the Striped Snakehead (*Channa striata*). *Helostoma temminkii* is not known from this lake, although recorded in nearby Lake Buyan (Green *et al*. 1978), *Ambassis* species are not known from this lake, and so these reads were likely runover from lab contamination of very high read counts observed in other lakes sequenced at the same time. *Barbodes lateristriga* and *Channa striata* reads were possibly contamination from the positive control used from tissue of these species, although both

Channa striata and a *Rasbora* species were recorded by Green *et al*. (1978) (*Barbodes laterstiga* was previously placed in the *Rasbora* genus). Some reads were removed based on their BLAST assignment, which may be actual eDNA from the lake, but may also be lab contamination. These were assigned to *Homo sapiens* (found in high abundance in all samples and also in low abundance in negative controls, and *Gallus gallus* (a known common contaminant of PCR reagents (Leonard *et al*. 2007), found in only one sample of the Lake Tamblingan data, not found in the Lake Tamblingan negative controls, but found in other lake samples and some of their negative controls).

4.5.5 COI marker data

The COI data, created using a primer pair (Leray *et al*. 2013) which amplifies a 313 bp fragment of the COI marker, was dominated by microfauna, meiofauna and microalgae. Consequently, this primer pair is not recommended for the detection of vertebrates from aquatic eDNA. There was minimal amplification of human DNA, but the BLAST hits for each OTU were of generally low quality matches. These data were useful however in examining patterns of community composition between points, and comparing OTU richness between points.

The highest quality BLAST assignment with a Query Cover and Identity of 100:99 was assigned to for OTU4 for the freshwater arthropod genus *Macrothrix* (a close relative of *Daphnia*), OTU363, assigned to the freshwater copepod genus *Moina*, and OTU113, assigned to the Class level of Polyarthra. The lowest quality BLAST assignment accepted was 89:73, matching to the Order Peronosporales (water moulds). The COI data created lower quality hits than the 12S and 16S data, with Query Cover ranging from 8 – 100 (before removing all assignments with a Query Cover of less than 55). There was very little amplification of human DNA, and very low read counts in negative controls before final filtering. However, there were some reads which were only found in negative controls, and some reads which were still present in negative controls, even with 3% background filtering, and so these OTUs were individually removed, (listed in Appendix 6). There were also reads found in the positive controls from the Anglesey Sea Zoo, which were found across the main samples. For example, OTU72, assigned to the genus *Macrothrix* was found in Lake Tamblingan samples as well as the positive control sample. Reads assigned to this OTU72 from this positive sample, ASZT2CN (from the Anglesey Sea Zoo in North Wales, 3,380 reads), was also in all samples from Lake Tamblingan apart from S5 and D2 (average 251 reads ± 66). This could

be possible as species within the *Macrothrix* genus are found worldwide, and individual populations are difficult to distinguish across countries (Neretina and Kotov, 2017).

Other OTUs were assigned to a variety of microfauna, meiofauna and microalgae mostly common in freshwater environments. OTUs 12 and 42 within the Cyclopidae family are copepods within the order Cyclopoida usually around 1 – 2 mm in size (Barnes, 1982). *Diaphanosoma excisum* (OTU 85) is a ctenopod within the family Sididae and order Cladocera of small crustaceans known as water fleas. OTUs 72, 4, 5, 88, 149 and 516 are also from this order of water fleas usually around 0.2 – 6 mm in size, as is OTU 363 assigned to *Moina*, a genus of water flea similar to *Daphnia* (Forró *et al*. 2008). OTUs 522 and 387 were assigned to Lepidoptera which includes moths and butterflies, within which many species are semiaquatic with the larval stage developing under water (Ward, 1992). OTU 18 was assigned to the order Opiliones, commonly known as harvestmen spiders, often found in large aggregations of many individuals near water (Pinto-da-Rocha *et al*. 2007). OTU 191 was assigned to to the Superorder of insects Holometabola (or Endopterygota), covering 850,000 possible species within butterflies, flies, fleas, bees, ants, and beetles (Beutel and Pohl, 2006). OTU 460 was assigned to the genus *Rhodotorula* which encompasses unicellular pigmented yeasts, a common environmental inhabitant in many environments including water (Wirth and Goldani, 2012). OTUs 129 and 89 assigned to the Family of Cryptomonadaceae, which are common freshwater algae (Barnes *et al*. 2009). OTU 415, 16 and 424 were assigned to Oomycetes, Oomycota and Peronosporales, a group of fungus-like eukaryotic microorganisms known as 'water moulds' which can use rhizoids to attach their thallus to the bed of stagnant or polluted water bodies (Sleigh, 1991). OTUs 44 and 102 were assigned to the diatom groups Thalassiosirales (Alverson, 2014) and Bacillatiophyta, a universally common group of microfauna, meiofauna and microalgae in aquatic habitats. OTU 596 (Chordariaceae), 47 (Dictyotaceae), 91 (Desmarestiales), and 107 (Ectocarpales) are brown algae. OTU 113 (Polyarthra), 170 (Flosculariaceae), and 1, 120, 10, 82 (Ploima) are rotifers, usually around 0.1 – 0.5 mm long, common in freshwater environments throughout the world (Segers, 2007)). OTU 90 was assigned to Heterokonts, which are a group encompassing algae, diatoms, water moulds and slime nets. These assignments therefore make ecological sense, as they encompass a broad range of freshwater associated organisms.

4.5.6 12S marker data

The 12S data, created using a primer pair (Valentini *et al*. 2016) which amplifies a ~ 120 bp fragment of the 12S marker, was composed almost entirely of vertebrate sequences, particularly teleost fish (as this primer pair was designed to do). Based on these data, I recommend this primer pair is for the detection of fish from aquatic eDNA. Valentini *et al*. (*2016*) compared the taxonomic coverage and resolution of their 'teleo' 12S primers with 12S primers designed by Riaz *et al*. (2011) and Thomsen *et al*. (2012b) and found that their own primers were more effective in amplifying teleost fish species (Valentini *et al*. 2016). Comparisons between other 12S primers would be useful, such as the MiFish primers designed by Miya *et al*. (2015).

There was significant amplification of human DNA, but the BLAST hits for each OTU were of generally high quality matches. The dominant fish species were those associated with aquaculture (Cichlids, Tilapia, Catfish and Carp), with remaining sequences assigned to invasive fish (Mosquitofish, Green swordtail and Guppy). This is not surprising based on the information available on the biodiversity and aquaculture fishery of this lake, which relies on regular restocking of a variety of aquaculture fish. Within the 12S data, there was a high quantity of reads assigned to the *Amphilophus* genus of Central American cichlids, found in all samples with 100% query cover and 98% identity match to both the cichlid fish *Amphilophus amarillo* and the Midas cichlid (*Amphilophus citrinellus*). Although fish from within this genus have not been recorded from Lake Tamblingan, they are a common aquaculture species in Bali, found in Lake Batur (Sentosa and Wijadi, 2012) and nearby Lake Beratan (Sentosa *et al*. 2013), and so it is likely that this fish has also been stocked into Lake Tamblingan for aquaculture purposes. There were also a high number of reads found assigned to the *Oreochromis* genus of Tilapia fish, found in 17/19 samples. Although the literature only mentions the Mozambique Tilapia (*Oreochromis mosambicus*) found at Lake Tamblingan, there were three unique OTUs assigned to tilapia fish: OTU20 and OTU12 (both *Oreochromis*) and OTU15 (*Pseudocrenilabrinae*). All matched with 100% BLAST Query Cover and Identity. The first possibility is that there is not enough variation within the 12S marker region used to distinguish between different tilapia species, and that three unique OTUs suggest there could be more than one species, possibly *O. mossambicus* and *O. niloticus*, both introduced locally for aquaculture, or even three different species. The second possibility is the OTU clustering approach split the same species into multiple OTUs. In a similar way, the 12S OTU58 assigned to Cyprinidae, but with Query Cover and Identity

of 100:91, and the 16S OTU36 assigned to Cyprinidae, but with a Query Cover and Identity of 100:94. These could be the result of sequencing artefacts, OTU clustering issues, or another Cyprinidae species than *Cyprinus carpio*. OTU19 matched with a Query Cover and Identity of 100:100, but only to a genus level assignment to *Osteochillus*, due to the voucher in the NCBI database only being described as 'Osteochilus sp'. This OTU is likely to belong to *Osteochilus vitattus* as recorded by Green *et al*. (1978). Neither the 12S gene or whole mitochondrion of *O. vitattus* is available in the NCBI database. The Walking Catfish (*Clarias batrachus*) and the African Sharptooth Catfish (*Clarias gariepinus*), both recorded from Lake Tamblingan (Green *et al*. 1978) were recovered from one sample point only (S9). This sample point was also the only point from which the widespread invasives the Common Carp (*Cyprinus carpio*) and the Western Mosquitofish (*Gambusia affinis*) were found. There were two other invasive fish found, one within the *Poecilia* genus (fish within this genus are known as Molly fish) and the Green Swordtail (*Xiphophorus hellerii*) in 2/19 and 6/19 sample points respectively. Although *X. helleri* was recovered from the 12S data, the Southern Platyfish (*X. maculatus*) recorded by Green *et al*. (1978) was not (both have 12S genes present in NCBI). The BLAST search resulted in a Query Cover and Identity of 100:100 matching to the top hit, *X. hellerii*. Other subsequent hits in the same BLAST result list were observed matching to *X. maculatus* with an identity of only 92, and so it appears that *X. hellerii* is the correct assignment out of the two for the observed OTU. Other studies record *X. hellerii* at nearby Lake Beratan (Sentosa *et al*. 2013) and Lake Buyan (Dahruddin *et al*. 2016). However, these two species can interbreed, producing fertile offspring (Schlosberg *et al*. 1949), and so it may be the case that either Green *et al*. (1978) wrongly identified *X. hellerii* as *X. maculatus,* or a hybrid was created, or alternatively, *X maculatus* was also present and not detected or no longer present at the lake, as this study was undertaken almost forty years ago. The sequence assigned to *Poecilia* (Query Cover : Identity = 100 : 95) most represents a species of Molly, the Guppy, *Poecilia reticulata*, recorded from Lake Tamblingan (Green *et al*. 1978) and other Balinese lakes including nearby Lake Buyan (Green *et al*. 1978), Beratan (Green *et al*. 1978; Whitten *et al*. 1996 Sentosa *et al*. 2013), and Lake Batur (Green *et al*. 1978; Sentosa and Wijaya, 2012; Budiasa *et al*. 2018).

4.5.7 16S marker data

The 16S data, created using a primer pair (Taylor *et al*. 1996) which amplifies a ~ 90 bp fragment of the 16S marker, was composed of both mammals and fish, although OTU richness was low overall, and samples were dominated by human contamination.

The 16S metabarcoding data recovered fewer OTUs, although it did recover a *Rasbora* (a genus of fish in the family Cyprinidae, native to freshwater habitats in South and Southeast Asia, and Southeast China) species with 99% BLAST Identity to all of *R. lateristriata*, *R. sumatrana* and *R. elegans*. It is therefore not possible to distinguish which of these species should be assigned to this OTU, although it is most likely to be *R. lateristriata* based on the literature. OTU7 was assigned to *Bos taurus* by MEGAN, although on inspection of the BLAST hits, this OTU also yielded a Query Cover and Identity of 100:100 to *B. primigenius* (Aurochs), *B. indicus* (Zebu Cattle), and *Phascolosoma esculenta* (a worm species commonly used for biological derivatives in biochemical research e.g. Wu *et al*. (2014)). *Bos taurus* DNA is a common contaminant in molecular pipelines (Leonard *et al*. 2007) due to the use of Bovine Serum Albumin (also known as BSA or "Fraction V") in reagents. As this OTU matched completely to all of these other species, it is likely that either the marker region cannot distinguish between these species, or these sequences are the result of errors on NCBI, or both. If this OTU was likely to be from local cattle, the local species *Bos javanicus domesticus* (with 16S genes available in NCBI) should have been observed, and so although there were no *Bos taurus* reads observed in the negative controls, this OTU was removed from the analysis. *Sus scrofa* (Wild Boar or Pig) was also recovered from the metabarcoding data, and is also a common laboratory reagent contaminant from gelatin used to purify *Taq* polymerases (Leonard *et al*. 2007). However, Wild Boar are native and fairly common in Bali (Whitten *et al*. 1996), and the top hits of the BLAST search were assigned *S. scrofus* from publications focusing on wild boar rather than *S. scrofus domesticus* (Domestic Pig). Additionally, no reads assigned to *S. scrofa* were observed in negative controls, and so this assignment was left in the analysis, as was the assignment to *Canis lupus* (Domestic Dog), a common sight around the Balinese lakes. None of the positive control tissue DNA (from the porpoise *Phocoena phocoena*), or the positive control eDNA were found in any of the other Lake Tamblingan samples, even though the *P. phocoena* DNA was present in very high read count (882,802) in the corresponding positive control cell of the OTU table, suggesting that the molecular workflow for the 16S dataset was less prone to 'bleeding' based contamination.

Before the removal of low quantity reads through bioinformatic filtering, the 16S marker also amplified the Convict Cichlid (*Amatitlania nigrofasciata*) known from the lake system around Lake Tamblingan, and so this highlights the trade off in eDNA metabarcoding of removing low quantity reads which may be generated from external contamination, and retaining low quantity reads which may be generated from real, low concentrations of target eDNA.

4.5.8 Limitations of the study and potential improvements

There was one sampling point (S9) which was more diverse than others, with 11 OTUs extracted from 12S data, 4 OTUs extracted from the 16S data, and 23 OTUs extracted from the COI data (totalling 38), compared with an overall average of 24 OTUs from other sampling sites. This sample point was particularly important for the 12S data, without which, five OTUs corresponding to four species and one genus would not have been detected. This may be due to the proximity of this sample point to the stream, where fish may prefer to reside. This highlights the need for many samples to be collected across a lake body to detect the resident biodiversity. This sample point was in the littoral zone close to the edge of the lake, and so it may be the case that eDNA accumulated at this shallow nearshore point. Based on the results presented here, it is unlikely that all biodiversity was detected using the described approach, and that a higher number of samples across the surface and a different depths, as well as a higher number of PCRs of separate extractions, would likely increase the number of species observed.

If this study were to be repeated, there are several changes to the sampling approach and molecular pipeline which may have resulted in better data. If sampling with remote sensing equipment for environmental variables (such as temperature) was possible, data richness of the samples taken at depth may have been improved. Another improvement would be to take samples at more fine scale distances apart, e.g. a grid system of every 100 m, and also to take samples deeper than 18 m. As the lake is permanently stratified into an oxygen depleted hypolimnion beginning at 29 m, it is expected that a more obvious community divide would occur beneath this depth if eDNA is fairly localised to its source individual. Additionally, if multiple temporal sampling events had been employed, the chance of detecting total biodiversity would increase. Samples may have shown higher quality read data if the filters had been immediately transported to the -80 °C freezer, rather than storing filters on ice or in a 4°C freezer overnight.

This study could have been improved further by keeping subsamples separate, to create true ecological replicates, processing these as individual extracts, and then as independent PCR replicates separately (although this would triple the molecular work load and cost of sequencing). Other eDNA studies have combined PCR replicates, as opposed to combining ecological replicates, although the disadvantage of this approach is the lack of ability to remove spurious sequences only found in one PCR. However, based on these results, if a sequence has entered the sample through lab contamination, it is likely to be of

such high read count that the source is obvious and these OTUs can be removed (such as the human DNA observed herein).

Field negative controls were added by removing a new Sterivex filter from the packet whilst at the side of the lake, and leaving it beside other samples whilst one sample was filtered. The field negative was then processed in the same way as test samples to test for contamination during the field sampling and transport phase. No water was filtered through the field negative controls, to remove the possibility of contamination from external water sources such as bottled water or distilled water taken from a laboratory. As Sterivex filter units are single use, there was no opportunity for contamination to occur between filtering of different samples, and so there was no need to filter clean water (e.g. distilled) as a negative control. However, other studies have used distilled water in their field negative controls (e.g Pilliod *et al*. 2013; Moyer *et al*. 2014), which may have generated a more faithful imitation of a test sample.

Another inadequacy of eDNA metabarcoding is the failure of short primers to discriminate between all species, as was observed herein. Incomplete barcoding information in public databases is a common issue for all metabarcoding studies, and particularly for under-studied areas such as Southeast Asia. These issues are further discussed in the General Discussion, as many topics refer to both Chapter 3 and Chapter 4.

A further limitation of this study was the interpretation of reads assigned to species which were used as positive controls. Several tissue samples were extracted from Malaysian and Indonesian fish to validate that primers were able to amplify these targets (see Universal Methods). To use up the remaining PCR wells, these extracts were added individually. A better option would have been to create a mock community using these extracts, by diluting them down to roughly that of the eDNA samples, and combining at equimolar ratios. This would have prevented such high read abundances sequenced in the positive control samples, and prevented the overspill contamination observed in other samples. It is impossible to know, therefore, whether these reads observed in the lake eDNA samples were genuine, or overspill from the positive controls. Based on this problem, I advise future studies to not sequence target species' DNA on the same sequencing run as eDNA samples which may contain these target species, and to dilute positive control samples down to roughly that of eDNA samples.

Temporal replication at the same site would also improve the understanding of eDNA information gained, confirming observed patterns. If repeated sampling was undertaken e.g. monthly across the year, seasonal patterns may be observed.

Another challenge related to this type of study is the tendency of Indonesian researchers to publish in Bahasa Indonesian, which (although of course natural and beneficial to Indonesian speakers) caused difficulty in establishing ecological information regarding site information and biodiversity present. For example, some studies only refer to 'Ikan Lele' or 'Lele Dumbo' (in Bahasa Indonesian) which only refers to the *Clarias* genus, not specific species, e.g. Negara *et al*. (2015).

Based on research (Spens *et al*. 2016, Chapter 2) generated after the sampling event for this study occurred, the breadth of eDNA information may have been enhanced by the use of a buffer such as Longmire's solution rather than freezing of filters after eDNA collection. The method of freezing samples was chosen to be logistically simpler in the field, and based on the Qubit results available at the time of sampling after undertaking a methods comparison study (Spens *et al*. 2016), freezing filters gave the highest overall concentration. However, based on Spens *et al*. (2016), and experience in the field with regards to the logistics of accessing suitable freezers and keeping samples cold enough during transport, the use of a buffer injected into the filter, such as Longmire's solution, is strongly advised. The use of a buffer, or dry storage on silica gel or beads, has also been recommended by several other studies mentioned in the Introduction (Chapter 1).

4.6 Conclusions

The number of reads per sample, and taxonomic community composition varied not only between surface samples spaced 500 m apart, but also between samples collected along the depth gradient only 2 m apart. This study demonstrates the necessity to undertake fine scale sampling at less than 500 m between points when targeting aquatic eDNA in tropical lacustrine environments. According to these data, if the objective of a sampling event is to record total resident biodiversity, and maximise the likelihood of detecting as many extant species as possible, 500 m distances or more between points is probably insufficient to capture all contained variety of unique eDNA barcodes. Sequences of COI, 12S and 16S barcodes recovered through metabarcoding of the aquatic eDNA samples of Lake Tamblingan exhibited spatial clustering, and unique community profiles at particular sites. This indicates that eDNA is not homogenously distributed across such a water body, and that the signal from a particular individual may be undetectable at distances less than 500m from

the source. This pattern has already been observed in previous studies discussed in the Introduction section of this chapter (Eichmiller *et al*. 2014; Yamamoto *et al*. 2016; Minamoto *et al*. 2017; Davidson *et al*. 2017; O'Donnell *et al*. 2016; Hänfling *et al*. 2016). Hänfling *et al*. 2016 found that most species were detected from shoreline samples of the lake, and they suggest eDNA could accumulate on the shoreline. The highest number of OTUs detected were from sample site S9, which could indicate that eDNA was accumulating at this shallow edge of the lake where this sample was collected close to the shoreline, but more intensive sampling would be necessary to verify this potential pattern.

The eDNA metabarcoding approach used here to survey extant biodiversity was able to amplify a broad range of life from microfauna, meiofauna and microalgae to large vertebrate fish. This has implications for food web analysis in which both the benthic food chain and the grazer food chain can be monitored to assess such patterns such as the impact of larger fish predators on lake eutrophication. Understanding the origin, state, transport and fate of eDNA in varying environments is essential if this technique is to be properly applied to ecological questions such as this. This aim will be better met by comprehensive, replicated sampling surveys across a range of species and habitats, drawing upon cross-disciplinary knowledge from e.g. microbiology and water quality monitoring. Based on the data herein, eDNA community composition is highly localised, and exhibits spatial variability at both the surface and depth gradients, meaning that future studies of tropical lake biodiversity should employ a sampling approach which covers as many spatial points as possible, focusing on shoreline sampling but including both surface samples and samples taken at depth.

References

Alverson, A.J., 2014. Timing marine–freshwater transitions in the diatom order Thalassiosirales. *Paleobiology*, 40(1), pp.91-101.

Anneville, O., Lasne, E., Guillard, J., Eckmann, R., Stockwell, J.D., Gillet, C. and Yule, D.L., 2015. Impact of fishing and stocking practices on coregonid diversity. *Food and Nutrition Sciences: FNS*, 6(11), pp.1045-1055.

Asian Development Bank. 2016. *Indonesia: Country water assessment*. Mandaluyong City, Philippines: Asian Development Bank, 2016.

Balasingham, K.D., Walter, R.P. & Heath, D.D., 2016. Residual eDNA detection sensitivity assessed by quantitative real-time PCR in a river ecosystem. *Molecular Ecology Resources*.

Bálint, M. *et al*., 2017. Twenty-five species of frogs in a liter of water : eDNA survey for exploring tropical frog diversity. , pp.0–36. *BioRxiv*.

Barnes, M.A. & Turner, C.R., 2016. The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, 17(1), pp.1–17.

Barnes, M.A., Turner, C.R., Jerde, C.L. *et al*. 2014. Environmental conditions influence eDNA persistence in aquatic systems. *Environ Science and Technology* 48:1819–1827.

Barnes, Richard Stephen Kent (2001). *The Invertebrates: A Synthesis.* Wiley-Blackwell. p. 41. ISBN 978-0-632-04761-1.).

Beutel, R.G. and Pohl, H., 2006. Endopterygote systematics–where do we stand and what is the goal (Hexapoda, Arthropoda)? Review. *Systematic Entomology*, 31(2), pp.202-219.

Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Douglas, W.Y. and De Bruyn, M., 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, 29(6), pp.358–367.

Boyer, F. *et al*., 2016. obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, 16(1).

Bulelengkab, 2013. PENEBARAN IKAN DI DANAU TAMBLINGAN DAN DANAU BUYAN KAB.BULELENG. (FISH GROUPS IN LAKE TAMBLINGAN AND LAKE BUYAN, BULELENG REGENCEY). Accessed at: https://diskan.bulelengkab.go.id/berita/penebaran-ikan-di-danau-tamblingan-dan-danau-buyan-kab-buleleng-1

Buxton, A.S. *et al*., 2017a. Is the detection of aquatic environmental DNA influenced by substrate type? H. Doi, ed. *PLOS ONE*, 12(8), p.e0183371.

Buxton, A.S. *et al*., 2017b. Seasonal variation in environmental DNA in relation to population size and environmental factors. *Scientific Reports*, 7, p.46294.

Bylemans, J. *et al*., 2016. An environmental DNA (eDNA) based method for monitoring spawning activity: a case study using the endangered Macquarie perch (*Macquaria australasica*). *Methods in Ecology and Evolution*.

Callahan, B.J., McMurdie, P.J. and Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME journal*, 11(12), p.2639.

Carim, K.J. *et al*., 2017. A noninvasive tool to assess the distribution of Pacific lamprey (Entosphenus tridentatus) in the Columbia River basin. *PLoS ONE*, 12(1).

Carim, K.J., McKelvey, K.S., Young, M.K., Wilcox, T.M., Schwartz, M.K. 2016. "A P Protocol for Collecting Environmental DNA Samples From Streams*". United States Department of Agriculture.*

Civade, R. *et al*., 2016. Spatial Representativeness of Environmental DNA Metabarcoding Signal for Fish Biodiversity Assessment in a Natural Freshwater System. *PLoS ONE*, 11(6).

Civade, R., Dejean, T., Valentini, A., Roset, N., Raymond, J.C., Bonin, A., Taberlet, P. and Pont, D., 2016. Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. *PloS one*, 11(6), p.e0157366.

Clarke, L.J., Beard, J.M., Swadling, K.M. and Deagle, B.E., 2017. Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. *Ecology and evolution*, 7(3), pp.873-883.

Cristescu, M.E. and Hebert, P.D. 2018. Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation. *Annual Review of Ecology, Evolution, and Systematics*, (0).

Darling, J.A., Mahon, A.R. 2011. From molecules to management: adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environmental Research* 111: 978–988.

Davison, P.I. *et al*., 2017. Application of environmental DNA analysis to inform invasive fish eradication operations. *The Science of Nature*, 104(3–4), p.35.

de Souza, L.S. *et al*., 2016. Environmental DNA (eDNA) Detection Probability Is Influenced by Seasonal Activity of Organisms H. Doi, ed. *PLOS ONE*, 11(10), p.e0165273.

de Ventura, L. *et al*., 2017. Tracing the quagga mussel invasion along the Rhine river system using eDNA markers: early detection and surveillance of invasive zebra and quagga mussels. *Management of Biological Invasions*, 8(1).

Deiner, K. and Altermatt, F., 2014. Transport distance of invertebrate environmental DNA in a natural river. *PLoS ONE*, *9*(2), p.e88786.

Deiner, K. *et al*., 2015. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, 183.

Deiner, K., *et al*. 2017. Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Molecular Ecology*.

Deiner, K., Fronhofer, E.A., Mächler, E., Walser, J.C. and Altermatt, F., 2016. Environmental DNA reveals that rivers are conveyer belts of biodiversity information. *Nature communications*, 7.

Dejean, T., Valentini, A., Duparc, A., Pellier-Cuit, S., Pompanon, F., Taberlet, P. and Miaud, C., 2011. Persistence of environmental DNA in freshwater ecosystems. *PloS one*, *6*(8), p.e23398.

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E. and 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*(4), pp.953-959.

Dunn, N., Priestley, V., Herraiz, A., Arnold, R. and Savolainen, V., 2017. Behavior and season affect crayfish detection and density inference using environmental DNA. *Ecology and Evolution*.

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(11).

Eichmiller, J.J., Best, S.E. & Sorensen, P.W., 2016. Effects of Temperature and Trophic State on Degradation of Environmental DNA in Lake Water. *Environmental Science and Technology*, 50(4).

Eidman, H.M., 1989. Exotic aquatic species introduction into Indonesia. Exotic aquatic organisms in Asia. *Asian Fisheries Society Special Publication*, 3, pp.57-62.

Eric O. Odada, Daniel O. Olago, Washington Ochola, Micheni Ntiba, Shem Wandiga, Nathan Gichuki and Helida Oyieke. 2015. PROCEEDINGS VOLUME I. 11TH WORLD LAKES CONFERENCE NAIROBI, KENYA, 31 OCTOBER TO 4TH NOVEMBER 2005. *Ministry of Eater and Irrigation International Lake Environment Committee.*

Evans, N.T. *et al*., 2017a. Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, p.cjfas-2016-0306.

Ficetola, G.F. *et al*., 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, 15(3), pp.543–556.

Ficetola, G.F., Miaud, C., Pompanon, F. and Taberlet, P., 2008. Species detection using environmental DNA from water samples. *Biology letters*, *4*(4), pp.423-425.

Forró, L., Korovchinsky, N.M., Kotov, A.A. and Petrusek, A., 2008. Global diversity of cladocerans (Cladocera; Crustacea) in freshwater. *Hydrobiologia*, 595(1), pp.177-184.

Fujiwara, A. *et al*., 2016. Use of environmental DNA to survey the distribution of an invasive submerged plant in ponds. *Freshwater Science*, 35(2).

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(2).

Gingera, T.D. *et al*., 2017. Environmental DNA as a detection tool for zebra mussels Dreissena polymorpha (Pallas, 1771) at the forefront of an invasion event in Lake Winnipeg, Manitoba, Canada. *Management of Biological Invasions*, 8

Goldberg, C.S. *et al*., 2013. Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshwater Science*, 32(3).

Goldberg, C.S. *et al*., 2016. Critical considerations for the application of environmental DNA methods to detect aquatic species. M. Gilbert, ed. *Methods in Ecology and Evolution*, 7(11), pp.1299–1307.

Goldberg, C.S., Pilliod, D.S., Arkle, R.S. and 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*(7), p.e22746.

Green, J., Corbet, S.A., Watts, E. and Lan, O.B., 1978. Ecological studies on Indonesian lakes. The montane lakes of Bali. *Journal of Zoology*, *186*(1), pp.15-38.

Hänfling, B. *et al*., 2016. Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology*

Ishige, T. *et al*., 2017. Tropical-forest mammals as detected by environmental DNA at natural saltlicks in Borneo. *Biological Conservation*.

Jane, S.F. *et al*., 2015. Distance, flow and PCR inhibition: EDNA dynamics in two headwater streams. *Molecular Ecology Resources*, 15(1).

Jerde, C.L. *et al*., 2016. Influence of Stream Bottom Substrate on Retention and Transport of Vertebrate Environmental DNA. *Environmental Science and Technology*, 50(16).

Jerde, C.L., Mahon, A.R., Chadderton, W.L. and Lodge, D.M., 2011. "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, 4(2), pp.150-157.

Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P.M., Woodcock, P., Edwards, F.A. and Larsen, T.H., 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology letters*, 16(10), pp.1245-1257.

Jo, T. *et al*., 2017. Rapid degradation of longer DNA fragments enables the improved estimation of distribution and biomass using environmental DNA. *Molecular Ecology Resources*.

Kalff, J., 2002. *Limnology: inland water ecosystems* (No. 504.45 KAL).

Kapoor, V., Elk, M., Toledo-Hernandez, C. and Santo Domingo, J.W., 2017. Analysis of human mitochondrial DNA sequences from fecally polluted environmental waters as a tool to study population diversity. *AIMS ENVIRONMENTAL SCIENCE*, 4(3), pp.443-455.

Kelly, R. P., Gallego, R., Jacobs-Palmer, E. 2018. The effect of tides on nearshore environmental DNA. *PeerJ* 6:e4521 https://doi.org/10.7717/peerj.4521

Kelly, R.P., Port, J.A., Yamahara, K.M. and Crowder, L.B., 2014. Using environmental DNA to census marine fishes in a large mesocosm. *PloS one*, *9*(1), p.e86175.

Keskin, E., 2014. Detection of invasive freshwater fish species using environmental DNA survey. *Biochemical Systematics and Ecology*.

Klymus, K.E. *et al*., 2015. Quantification of eDNA shedding rates from invasive bighead carp Hypophthalmichthys nobilis and silver carp *Hypophthalmichthys molitrix*. *Biological Conservation,* 183.

Lacoursière-Roussel, A. *et al*., 2016a. Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. *Journal of Applied Ecology*, 53(4).

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(6).

Lance, R.F. *et al*., 2017. Experimental observations on the decay of environmental DNA from bighead and silver carps. *Management of Biological Invasions*, 8.

Laramie, M.B., Pilliod, D.S., Goldberg, C.S., and Strickler, K.M. 2015. "Environmental DNA Sampling Protocol – Filtering Water to Capture DNA from Aquatic Organisms". *U.S. Geological Survey*.

Larson, E.R. *et al*., 2017. Environmental DNA (eDNA) detects the invasive crayfishes *Orconectes rusticus* and Pacifastacus leniusculus in large lakes of North America. *Hydrobiologia*, pp.1–13.

Lawson Handley, L., 2015. How will the 'molecular revolution' contribute to biological recording?. Biological *Journal of the Linnean Society*, 115(3), pp.750-766.

Lehmusluoto, P., Machbub, B., Terangna, N., Rusmiputro, S., Achmad, F., Boer, L., Brahmana, S.S., Priadi, B., Setiadji, B., Sayuman, O. and Margana, A., 1997. National inventory of the major lakes and reservoirs in Indonesia. *Expedition Indodanau Technical Report, Edita Oy*.

Leonard, J.A., Shanks, O., Hofreiter, M., Kreuz, E., Hodges, L., Ream, W., Wayne, R.K. and Fleischer, R.C., 2007. Animal DNA in PCR reagents plagues ancient DNA research. *Journal of Archaeological Science*, 34(9), pp.1361-1366.

Leray, M. *et al*., 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in zoology*, 10(1), p.34.

Li, J., Lawson Handley, L.J., Read, D.S. and Hänfling, B., 2018. The effect of filtration method on the efficiency of environmental DNA capture and quantification via metabarcoding. *Molecular ecology resources.*

Likens, G.E. ed., 2010. Lake ecosystem ecology: A global perspective. *Academic Press*.

Lim, N.K.M. *et al*., 2016. Next-generation freshwater bioassessment: eDNA metabarcoding with a conserved metazoan primer reveals species-rich and reservoir-specific communities. *Royal Society Open Science*, 3(11).

Lodge, D.M., Turner, C.R., Jerde, C.L., Barnes, M.A., Chadderton, L., Egan, S.P., Feder, J.L., Mahon, A.R. and Pfrender, M.E., 2012. Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology*, *21*(11), pp.2555-2558.

Maghfiroh, M., Dianto, A., Jasalesmana, T., Melati, I., Samir, O., Kurniawan, R. 2016. Lake Ecosystem Health and Its Resilience: Diversity and Risks of Extinction. *PROCEEDINGS of the 16th World Lake Conference.*

Maruyama, A. *et al*., 2014. The release rate of environmental DNA from juvenile and adult fish. *PLoS ONE*, 9(12).

Matsui, K., Honjo, M., Kawabata, Z. 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

Minamoto, T. *et al*., 2017. Environmental DNA reflects spatial and temporal jellyfish distribution H. Doi, ed. *PLOS ONE*, 12(2), p.e0173073.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J.Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H. and Kondoh, M., 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society open science*, 2(7), p.150088.

Moyer, G.R. *et al*., 2014. Assessing environmental DNA detection in controlled lentic systems. *PLoS ONE*, 9(7).

Nathan, L.M. *et al*., 2014. Quantifying environmental DNA signals for aquatic invasive species across multiple detection platforms. *Environmental Science and Technology*, 48(21).

Negara, I.K.W., Marsoedi, M. and Susilo, E., 2015. Strategi Pengembangan Budidaya Lele Dumbo Clarias SP. Melalui Program Pengembangan USAha Mina Pedesaan Perikanan Budidaya Di Kabupaten Buleleng (**Aquaculture Development of Catfish Clarias SP. Through Fisheries Business Development in Village on Fish Aquac**. *Jurnal Manusia dan Lingkungan*, 22(3), pp.365-371.

Neretina, A.N. and Kotov, A.A., 2017. Diversity and distribution of the Macrothrix paulensis species group (Crustacea: Cladocera: Macrothricidae) in the tropics: what can we learn from the morphological data*?*. *In Annales de Limnologie-International Journal of Limnology* (Vol. 53, pp. 425-465). EDP Sciences.

New England BioLabs inc. NEBNext® DNA Library Prep Master Mix Set for 454™ Instruction Manual.

O'Donnell, J.L., Kelly, R.P., Shelton, A.O., Samhouri, J.F., Lowell, N.C. and Williams, G.D., 2017. Spatial distribution of environmental DNA in a nearshore marine habitat. *PeerJ*, 5, p.e3044.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H. and Oksanen, M.J., 2013. Package 'vegan'. *Community ecology package, version*, 2(9).

Pedersen, M.W., Overballe-Petersen, S., Ermini, L., Der Sarkissian, C., Haile, J., Hellstrom, M., Spens, J., Thomsen, P.F., Bohmann, K., Cappellini, E. and Schnell, I.B., 2015. Ancient and modern environmental DNA. *Philosophical Transactions, Royal Society B*, 370(1660), p.20130383.

Pfleger, M.O. *et al*., 2016. Saving the doomed: Using eDNA to aid in detection of rare sturgeon for conservation (Acipenseridae). *Global Ecology and Conservation*, 8.

Piaggio, A.J., Engeman, R.M., Hopken, M.W., Humphrey, J.S., Keacher, K.L., Bruce, W.E. and Avery, M.L., 2014. Detecting an elusive invasive species: a diagnostic PCR to detect Burmese python in Florida

waters and an assessment of persistence of environmental DNA. *Molecular Ecology Resources*, *14*(2), pp.374-380.

Pilliod, D.S. *et al*., 2013. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, 70(8).

Pilliod, D.S. *et al*., 2014. Factors influencing detection of eDNA from a stream-dwelling amphibian. *Molecular Ecology Resources*, 14(1).

Pinto-da-Rocha, R., Machado, G. and Giribet, G. eds., 2007. Harvestmen: the biology of Opiliones. *Harvard University Press.*

Porazinska, D.L., GIBLIN-DAVIS, R.M., Esquivel, A., Powers, T.O., Sung, W.A.Y. and Thomas, W.K., 2010. Ecometagenetics confirm high tropical rainforest nematode diversity. *Molecular Ecology*, 19(24), pp.5521-5530.

Port, J.A., O'Donnell, J.L., Romero-Maraccini, O.C., Leary, P.R., Litvin, S.Y., Nickols, K.J., Yamahara, K.M. and Kelly, R.P., 2016. Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*, 25(2), pp.527-541.

Puillandre, N. *et al*., 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, 21(8), pp.1864–1877.

Razgour, O., Clare, E.L., Zeale, M.R., Hanmer, J., Schnell, I.B., Rasmussen, M., Gilbert, T.P. and Jones, G., 2011. High-throughput sequencing offers insight into mechanisms of resource partitioning in cryptic bat species. *Ecology and Evolution*, 1(4), pp.556-570.

Rees, H.C. *et al*., 2014. The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, 51(5).

Rees, H.C. *et al*., 2015. Applications and limitations of measuring environmental DNA as indicators of the presence of aquatic animals. *Journal of Applied Ecology*, 52(4).

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P. and Coissac, E., 2011. ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, 39(21), pp.e145-e145.

Robert D. Barnes (1982). *Invertebrate Zoology*. Philadelphia, Pennsylvania: Holt-Saunders International. pp. 683–692. ISBN 0-03-056747-5.).

Robson, H.L., Noble, T.H., Saunders, R.J., Robson, S.K., Burrows, D.W. and 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(4), pp.922-932.

Royle JA, Link WA (2006) Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, **87**, 835–841.

Sassoubre, L.M. *et al*., 2016. Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine Fish. *Environmental Science and Technology*, 50(19).

Schlosberg, H., Duncan, M.C. and Daitch, B.H., 1949. Mating behavior of two live-bearing fish, Xiphophorus hellerii and Platypoecilus maculatus. *Physiological zoology*, *22*(2), pp.148-161.

Schnell, I.B., Bohmann, K. and Gilbert, M.T.P., 2015. Tag jumps illuminated–reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, 15(6), pp.1289-1303.

Schnell, I.B., Thomsen, P.F., Wilkinson, N., Rasmussen, M., Jensen, L.R., Willerslev, E., Bertelsen, M.F. and Gilbert, M.T.P., 2012. Screening mammal biodiversity using DNA from leeches. *Current biology*, 22(8), pp.R262-R263.

Schubert, M., Lindgreen, S. & Orlando, L., 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging Findings Background. *BMC Res Notes*, 9.

Segers, H., 2007. Annotated checklist of the rotifers (Phylum Rotifera), with notes on nomenclature, taxonomy and distribution. *Zootaxa.*

Sentosa, A.A. and Wijaya, D., PEMANFAATAN MAKANAN ALAMI OLEH IKAN-IKAN DOMINAN DI DANAU BATUR, PROVINSI BALI. *Seminar Nasional Tahunan IX Hasil Penelitian Perikanan dan Kelautan, 14 Juli 2012* (**UTILIZATION OF NATURAL FOOD BY DOMINANT FISH IN LAKE BATUR, BALI PROVINCE**.).

Sentosa, A.A., Wijaya, D. and Tjahjo, D.W.H., 2013. Kajian risiko keberadaan ikan-ikan introduksi di Danau Beratan, Bali. In *Prosiding Forum Nasional Pemulihan dan Konservasi Sumberdaya Ikan IV* (pp. 1-16). (**Risk assessment of the presence of introduced fish in Lake Beratan, Bali. In Instruction of the National Recovery and Cons Forum Conservation of Fish Resources IV**)

Seymour, M., Deiner, K. & Altermatt, F., 2016. Scale and scope matter when explaining varying patterns of community diversity in riverine metacommunities. *Basic and Applied Ecology*, 17(2).

Seymour, M., Durance, I., Cosby, B.J., Ransom-Jones, E., Deiner, K., Ormerod, S.J., Colbourne, J.K., Wilgar, G., Carvalho, G.R., de Bruyn, M. and Edwards, F., Gregory Wilgar, Gary R. Carvalho, Mark de Bruyn, François Edwards, Bridget A. Emmett, Holly M. Bik & Simon Creer. 2018. Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms. *Communications Biology*, 1(1), p.4.

Shaw, J.L., Clarke, L.J., Wedderburn, S.D., Barnes, T.C., Weyrich, L.S. and Cooper, A. 2016. Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biological Conservation*, 197, pp.131–138.

Sigsgaard, E.E. *et al*., 2017. Seawater environmental DNA reflects seasonality of a coastal fish community. *Marine Biology*, 164(6), p.128.

Sleigh, M.A., 1991. *Protozoa and other protists*. CUP Archive.

Spens, J. *et al*., 2016. Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods in Ecology and Evolution*, 8(5), pp.635-645..

Stoeckle, M.Y. *et al*., 2017. Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary H. Doi, ed. *PLOS ONE*, 12(4), p.e0175186.

Taberlet, P., Coissac, E., Hajibabaei, M. and Rieseberg, L.H., 2012a. Environmental DNA. *Molecular ecology*, *21*(8), pp.1789-1793.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. and Willerslev, E., 2012b. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular ecology*, *21*(8), pp.2045-2050.

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H., Kawabata, Z. 2012. Estimation of fish biomass using environmental DNA. *PLoS One* 7: e35868.

Taylor, P.G., 1996. Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. *Molecular biology and evolution*, 13(1), pp.283-285.

Thomsen, P., Kielgast, J.O.S., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., Orlando, L. and Willerslev, E., 2012a. Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular ecology*, *21*(11), pp.2565-2573.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Møller, P.R., Rasmussen, M. and Willerslev, E., 2012b. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS one*, *7*(8), p.e41732.

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

Torresdal, J.D., Farrell, A.D. and Goldberg, C.S., 2017. Environmental DNA detection of the golden tree frog (Phytotriades auratus) in bromeliads. *PloS one*, 12(1), p.e0168787.

Tréguier, A., Paillisson, J.M., Dejean, T., Valentini, A., Schlaepfer, M.A., Roussel, J.M., 2014. Environmental DNA surveillance for invertebrate species: advantages and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. *Journal of Applied Ecology*. 51:871–879. doi: 10.1111/ 1365-2664.12262.

Tsuji, S. *et al*., 2017a. Water temperature-dependent degradation of environmental DNA and its relation to bacterial abundance. *Plos One*, 12(4).

Tsuji, S., Yamanaka, H. & Minamoto, T., 2017b. Effects of water pH and proteinase K treatment on the yield of environmental DNA from water samples. *Limnology*, 18(1).

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.

Ushio, M. *et al*., 2017. Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. *bioRxiv*.

Valdez-Moreno, M., Ivanova, N.V., Elias-Gutierrez, M., Pedersen, S.L., Bessonov, K. and Hebert, P.D., 2018. Using eDNA to biomonitor the fish community in a tropical oligotrophic lake. *bioRxiv*, p.375089.

Valentini, Alice, Taberlet, Pierre, Miaud, Claude, Civade, Raphaël, Herder, Jelger, Thomsen, Philip Francis, Bellemain, E., 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, 25, pp.929–942.

Vervoort, M.T.W., Vonk, J.A., Mooijman, P.J.W. *et al*. 2012. SSU ribosomal DNA-based monitoring of nematode assemblages reveals distinct seasonal fluctuations within evolutionary heterogeneous feeding guilds. *PLoS ONE* 7:e47555.

Vo, A.T.E. & Jedlicka, J.A., 2014. Protocols for metagenomic DNA extraction and Illumina amplicon library preparation for faecal and swab samples. *Molecular Ecology Resources*.

Ward, J.V., 1992. *Aquatic insect ecology*. 1. Ecology and habitat. John Wiley & Sons, Inc..

Whitten, T., Soeriaatmadja, R.E. and Afiff, S.A., 1996. Ecology of Java & Bali (Vol. 2). *Oxford University Press.*

Wickham, H., 2016. ggplot2: elegant graphics for data analysis. *Springer*.

Williams, P. 2013. "GCN eDNA protocol". PondNet. *Freshwater Habitats Trust*.

Wirth, F. and Goldani, L.Z., 2012. Epidemiology of Rhodotorula: an emerging pathogen. *Interdisciplinary perspectives on infectious diseases*, 2012.

Wu, Y., Fang, M., Du, L., Wu, H., Liu, Y., Guo, M., Xie, J. and Wei, D., 2014. The nutritional composition and anti-hypertensive activity on spontaneously hypertensive rats of sipuncula Phascolosoma esculenta. *Food & function,* 5(9), pp.2317-2323.

Yamamoto, S., Minami, K., Fukaya, K., Takahashi, K., Sawada, H., Murakami, H., Tsuji, S., Hashizume, H., Kubonaga, S., Horiuchi, T. and Hongo, M., 2016. Environmental DNA as a 'snapshot'of fish distribution: A case study of Japanese jack mackerel in Maizuru Bay, Sea of Japan. *PloS one*, 11(3), p.e0149786.

Yoccoz, N.G., 2012a. The future of environmental DNA in ecology. *Molecular Ecology*, 21(8), pp.2031–2038.

Zepeda-Mendoza, M.L. *et al*., 2016. DAMe: a toolkit for the initial processing of datasets with PCR replicates of double-tagged amplicons for DNA metabarcoding analyses. BMC research notes, 9(1), p.255.

Online References

Bali Travel News. 2016. Fingerling released into Lake Buyan and Lake Tamblingan. Activities. Bali Travel News. http://bali-travelnews.com/2016/07/08/fingerling-released-into-lake-buyan-and-lake-tamblingan/

Lake Lubbers. 2018. Lake Buyan and Lake Tamblingan-Bali's Twin Lakes, Indonesia. https://www.lakelubbers.com/lake-buyan-and--lake-tamblinganbalis-twin-lakes-2453/

Siriwardena, S. 2010. Gambusia affinis (western mosquitofish). Invasive Species Compendium. CABI Centre for Agriculture and Bioscience International. Accessed at:

https://www.cabi.org/isc/datasheet/82079Weatherbase, 2018. BALI, INDONESIA. http://www.weatherbase.com/weather/weatherall.php3?s=3279&cityname=Bali%2C+Bali%2C+Indonesia&units=

Weather Underground, 2018. Weather History for WRRR - July, 2015.

https://www.wunderground.com/history/airport/WRRR/2015/6/4/MonthlyHistory.html?req_city=Tamblingan&req_state=BA&req_statename=Indonesia&reqdb.zip=00000&reqdb.magic=1210&reqdb.wmo=97230

Chapter 5

Assessing the freshwater biodiversity of the lakes of the Malay Archipelago using eDNA metabarcoding

5.1 Abstract

The Malay Archipelago contains some of the highest biodiversity in the world and has particularly high freshwater ichthyofaunal diversity. In this study, the use of aquatic eDNA metabarcoding to detect extant biodiversity from Indonesian and Malaysian lakes was tested for the first time. Water was collected along transects of each lake, and filtered to capture eDNA. A range of fish, mammals, amphibians, invertebrates, microfauna, meiofauna and microalgae were detected with high confidence, including many native freshwater and freshwater associated species, from common aquaculture fish to a rare primate thought locally extinct. Nearly all species detected were known from the literature or could be explained using biological knowledge of the area. The biodiversity and ecological communities from different lakes and regions showed differences in species richness and community composition, and also with respect to habitat variables including altitude, trophic productivity, area, and maximum depth. The turbidity of some of these highly disturbed Southeast Asian lakes proved challenging for filtering, however this study was an overall success in demonstrating the feasibility of eDNA monitoring in Southeast Asian freshwater habitats. Although improvements have been identified here when employing this type of aquatic eDNA metabarcoding, this study proves the potential for this approach in monitoring aquatic-associated species including invasive fish and molluscs from biodiversity hotspots such as the mega-diverse Malay Archipelago.

5.2 Introduction

5.2.1 Freshwater biodiversity of the Malay Archipelago

The Malay Archipelago is the largest archipelagic area in the world, constituting 25,000 islands covering six countries including Brunei Darussalam, Indonesia, Malaysia, the Philippines, Singapore, and Timor-Leste. This archipelago stretches 6,100 km along the equator and 3,500 km north to south (Encyclopaedia Britannica, 2018). For the purpose of this study, Indonesia and Malaysia (which have the highest fish species richness of these countries, shown in Figure 5.1.1 below) are the focus. Of particular interest within the Malay Archipelago is the infamous Wallace Line, where the two continents of Asia and Australia meet, which runs between Sulawesi and Borneo and Bali and Lombok. East of Wallace's Line, primary freshwater fishes such as cyprinids do not naturally exist, although a large number of species have now been introduced (Coates, 1985; Coates, 2002). In tropical Asian lakes, fish species richness is mostly predicted by lake area rather than other variables such as temperature, pH and primary productivity which predict fish species richness in temperate lakes (Amarasinghe and Welcomme, 2002). Indonesia is designated as one of the megadiverse countries of the world, behind only Brazil (Collen *et al.* 2014), with an estimated 4,000 fish species, at least 1,000 of which are freshwater (Suwelo, 2004), figures which are likely underestimated as new species are being discovered at a rate of around 200 species per year (Nelson, 1994). It is likely that actually more than 1,300 freshwater fish species reside in Indonesia, with roughly 798 species from Sundaland, 68 from Wallacea, and 58 from Sahul zones of the country (Kartamihardja, 2015), numbers of which are shown in comparison with other countries in Figure 5.1 below.

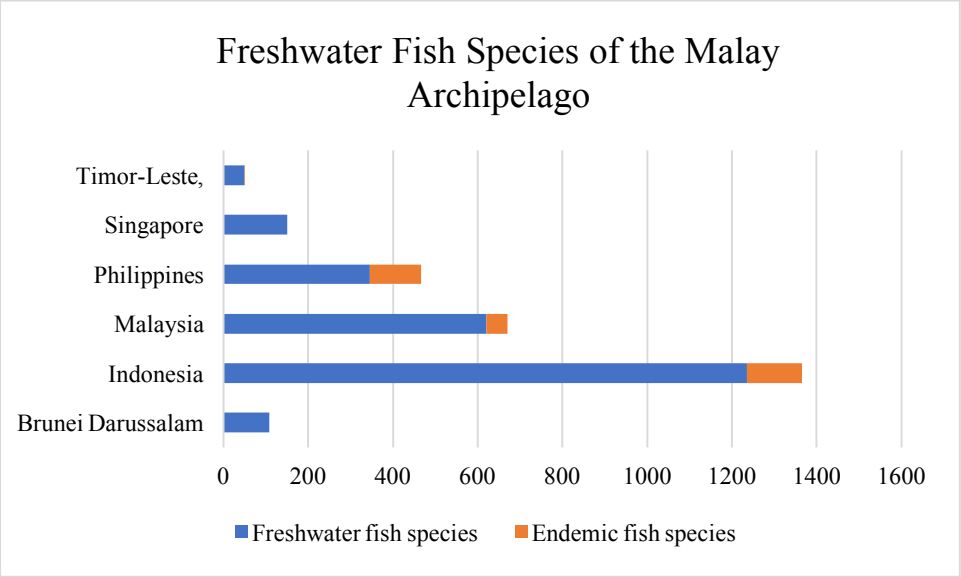


Figure 5.1 Freshwater Fish Species of the Malay Archipelago. Species numbers from fishbase.org

5.2.2 Importance of Lakes in Malay Archipelago

There is a lack of natural freshwater lakes, formed by glaciers, volcanic activities and tectonic movements (as opposed to manmade reservoir lakes) in tropical Asia, with the majority being found in Indonesia and the Philippines. Many of the existing lakes are however, extremely important in terms of fisheries and aquatic biodiversity (De Silva, 2010; Amarasinghe and De Silva, 2015). There are 840 major and 736 small lakes, as well as 162 major reservoirs and 1,341 small reservoirs in Indonesia (Kartamihardja, 2015). Indonesia has three types of reservoirs, 1. field reservoirs (community authority, for water supply), 2. irrigation reservoirs (local government authority, for agriculture) and 3. multipurpose reservoirs (central government authority, for e.g. flood control, hydroelectric power, irrigation and water supply) (Hardjamulia and Suwigno, 1988). The creation of such reservoirs usually requires placing a dam across a river to artificially create a reservoir lake. In areas such as Peninsular Malaysia where water resources are heavily impounded, many people have been displaced after the creation of such dams and have adopted cage fish farming as an alternative livelihood (De Silva, 2010).

Asia contributes 69% of the world's inland capture fisheries and aquaculture production, increasing by 43% from 2004 to 2010 with global growth almost completely attributable to Asia (Amarasinghe and De Silva, 2015). The total fisheries production volume of Indonesia alone was around 18.8 million metric tons in 2012, accounting for 47% of Southeast Asian fisheries production, 57% of which came from aquaculture (Kartamihardja, 2015). As rice is such an important crop for Asian countries, integrated rice-fish culture is practised in many countries including Indonesia (De Silva, 2010). Culture-based-fisheries (CBF) involve the release of hatchery-produced seeds and juveniles into water bodies, where they consume natural foods until reaching market size (Kartamihardja, 2015).

Fish contribute a plethora of fundamental ecosystem services, including regulating, linking, and demand-derived services. Regulating ecosystem services include the regulation of food web dynamics, nutrients, biodiversity, ecosystem resilience, redistribution of bottom substrates, carbon flux and sediment processes. Linking ecosystem services include linkage within aquatic ecosystems, between aquatic and terrestrial ecosystems, transport of nutrients, inorganic compounds and energy. Demand-derived ecosystem services include the provision of cultural services, food, medicine, disease control, aquatic plant control, reduction of waste, recreational activities, assessment of ecosystem stress and resilience, revealing evolutionary tracks and providing scientific and educational information (Holmlund and Hammer, 1999).

5.2.3 Threats to biodiversity within the lakes of the Malay Archipelago

The main threats to lacustrine freshwater biodiversity are water pollution, flow modification, habitat degradation, over exploitation, species invasions and environmental change, which are particularly prevalent in the Malay Archipelago (further explained in Appendix 6 'How can we conserve the imperilled freshwater ecosystems of Southeast Asia?'). Multipurpose reservoir construction has accelerated over the latter half of the 20th century, mostly for hydroelectric power, and agricultural irrigation, with fisheries becoming a significant secondary use of these impounded waters (Amarasinghe and De Silva, 2015). Asian lacustrine fisheries have a significant impact on rural livelihoods and nutrition of rural people, but have not received adequate policy control, research, development, or technology (De Silva, 2010; Amarasinghe and De Silva, 2015).

The inland waters of Indonesia are under one Fisheries Management Area and can be used for fisheries and aquaculture development. Indonesia is one of the world's top aquaculture producers (Amarasinghe and De Silva, 2015). Stock enhancement and CBF of inland waters are promoted by Southeast Asian countries, and particularly Indonesia, for fish production, food security, income for fishers and human wellbeing (Kartamihardja, 2015; Amarasinghe and De Silva, 2015). Of the 840 major lakes of Indonesia, there are 28 key lakes for stock enhancement and restocking, including Lake Laut Tawar, Lake Toba, Lake Singkarak, Lake Semayang, Lake Melintang, Lake Matano, and Lake Batur, all sampled in this study (Kartamihardja, 2015).

In Peninsular Malaysia, water resources are heavily impounded, and a relatively large number of ornamental fishes are produced for the export trade (Coates, 2002). Malaysia has a modest open-water stocking programme of mostly Silver Barb (*Barbonymus gonionotus*), Common Carp (*Cyprinus carpio*), Giant Freshwater Prawn (*Macrobrachium rosenbergii*), Red Tilapia (*Oreochromis niloticus* red-hybrid) and River Catfish (*Pangasius* sp.) (Coates, 2002).

Stocking as recompense for decreasing fish populations often creates artificial systems which are dependent upon a constant input of reared fish and may disguise ecological patterns which consequently weaken the implementation of conservation management (Holmlund and Hammer, 1999). In many cases, stocking can also cause depletion of other economically valuable species, changes in nutrient balances, or biodiversity decline (Holmlund and Hammer, 1999). With this type of stocking-based fishery in lakes and reservoirs, increased fishing pressure, open access and unregulated fisheries are a problem and are often associated with biologically incompatible reservoir water level management (Petr 1995). Aquaculture in the form of cage culture or floating cages within

lakes can increase phosphates and nitrates and lead to eutrophication (Pratiwi *et al*. 2016). Many environmental problems are associated with aquaculture in lakes and reservoirs, especially due to inadvertent expansion of cage culture which when practices are over-intensified, can cause deterioration of water quality, resulting in fish kills (Abery *et al*. 2005). Other sources of nutrient loading come from agriculture and residential wastes, and so are generally an indicator of human impact. Indonesian freshwaters suffer from a host of fishing related threats, such as fishing by tipping large quantities of DDT or rotenone into the water, electric fishing and underwater explosives. Additional pollutants from agricultural pesticides and mineral extraction are a problem, particularly for lakes that are small or slow-flowing (Whitten *et al*. 1996). Due to these types of threats to lacustrine ecosystems, there are fifteen national priority lakes designated by the Indonesian government for rehabilitation in Indonesia, listed here with lakes used in this study in bold: in Sumatra (**Lake Toba**, Maninjau, **Singkarak**, Kerinci), in Sulawesi (Tondano, Limboto, Poso, Tempe, **Matano**), in Kalimantan (Mahakam **Semayang-Melintang**-Jempang, Sentarum), in Papua (Sentani), in Banten (Rawa Danau), in Bali (**Batur**) and in Central Java (**Rawa Pening**) (Haryani, 2016).

5.2.4 Environmental DNA for biodiversity monitoring of tropical lakes

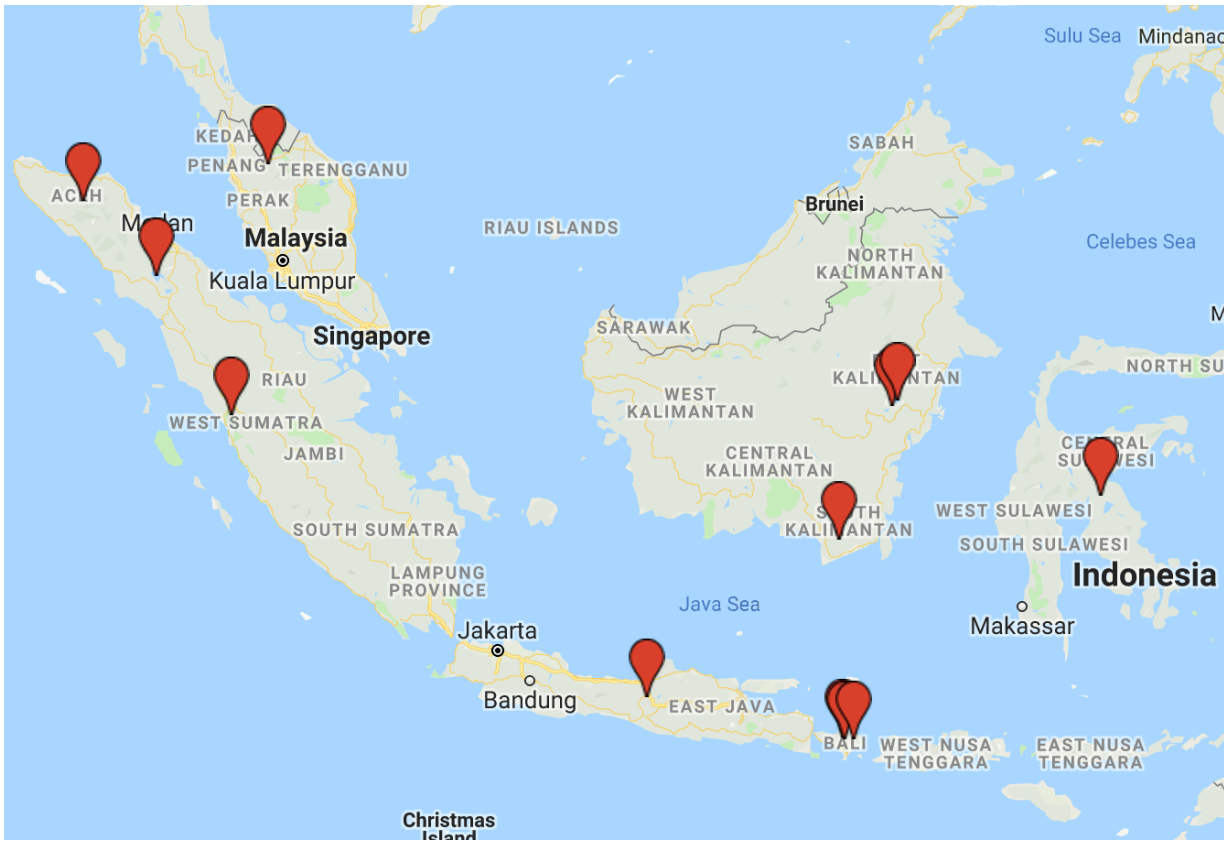
Freshwater fauna is particularly sensitive to environmental change and disruption (Brander 2007; Dudgeon 2010), and consequently management agencies often use the status of regional fish and amphibian biodiversity as ecosystem health indicators to prioritize and assess management strategies (Sala *et al*. 2005; Xenopoulos *et al*. 2005; Abell *et al.* 2008; Giller *et al*. 2004). This relies on accurate population assessments in the field, including species richness, diversity, distribution and abundance. Conventional methods of aquatic bioassessment have depended upon catching individual organisms such as fish, using gill nets, long-lining, traps, acoustic monitoring, baited remote underwater video (BRUV), underwater visual census (UVC) and fisheries-dependent population surveys or electrofishing (Murphy & Willis 1996; Bonar *et al*. 2009). These methods are destructive, labour intensive, expensive, require taxonomic expertise, have bias, and cannot always give a complete picture of biodiversity due to inefficiencies of sampling, meaning that false negatives may arise concerning rare or elusive species (Lodge *et al*. 2012; Argillier *et al*. 2013; Kubečka *et al*. 2009; Bayley & Peterson 2001; Mackenzie & Royle 2005). Conservation management and ecological research can therefore be hindered when using these conventional methods if changes in biodiversity cannot be rapidly assessed. One of the major priorities of the

International Decade for Action 'Water for Life' programme which ran from 2005 to 2015 is freshwater biodiversity conservation (UN, 2015). To understand the resource potential of lakes, baseline data is needed to assess biological and non-biological natural resources, with implications for the sustainable development of fisheries, tourism and conservation (Restu *et al*. 2016).

In large lake ecosystems, established methods are currently inadequate to fulfil legislative obligations such as the EC Water Framework (European Communities 2000). There is a need for more concerted efforts to monitor fish catches in Southeast Asia, where fish production and inland fishery statistics do not even differentiate between the type of water body, even though the majority of inland fisheries for food fish production occur in lacustrine waters in Asian countries (Coates, 2002; Amarasinghe and De Silva, 2015). To understand the effects of anthropogenic impact such as mining pollutants on individual species and the ecosystem as a whole, it is first necessary to understand what species are present in a community, and consequently, which species may come into contact with such impacts. The implementation of eDNA sampling could provide a viable solution to these questions, allowing environmental biodiversity monitoring to inform regulators or managers of conservation priorities, fisheries population patterns and the spread of invasive species.

Most aquatic eDNA studies have focused on the detection of single species using species-specific markers, and only recently has the detection of species communities based on eDNA metabarcoding become more common. Hänfling *et al*. (2016) detected 14/16 species recorded at Lake Windermere, compared with 4/16 detected by gill net surveys. Valentini *et al*. (2016) detected amphibian and fish species from lakes and ponds using eDNA where conventional surveys proved less successful, with an overall detection probability of 0.97 and 0.58 respectively. Keskin *et al*. (2016) detected twenty-three species of fish from a Turkish lake, five of which were reported for the first time. Civade *et al*. (2016) using the same methods as Valentini *et al*. (2016), detected 21/26 taxa from three eDNA metabarcoding samples, compared to 22/26 from seven cumulated traditional sampling surveys of ponds, lakes and rivers. Evans *et al*. (2017) detected all of the fish species detected by traditional sampling from a pond in the USA and eleven additional species not detected using traditional sampling. However, results varied depending on the bioinformatic stringency employed. Most recently, Valdez-Moreno *et al*. (2018) detected seventy-five species of vertebrates including forty-seven fishes, fifteen birds, seven mammals, five reptiles, and one amphibian from a Mexican lake. As discussed in Chapter 3,

although there have been metabarcoding studies from the tropics from iDNA (Schnell *et al*. 2012), soil eDNA (Porazinkska *et al*. 2010; Yoccoz *et al*. 2012), and marine water targeting microbiota (e.g Rusch *et al*. 2007), there have been few aquatic eDNA studies targeting macrobial life. There have been tropical, aquatic, macrobial eDNA studies focusing on single-species (Piaggio *et al*. 2014; Robson *et al*. 2016), marine systems (Bakker *et al*. 2017) or single aquatic habitats with high biodiversity traffic (Ishige *et al*. 2017) and human haplotype variation (Kapoor *et al*. 2017). There are a number of very recent studies using eDNA metabarcoding from natural freshwater bodies: Bálint *et al*. (2017) detected twenty-five species of frog from ponds in Bolivia, and Cilleros *et al*. (2018) detected 132 fish species from Guianese sites. However, to our knowledge, this is the first aquatic eDNA metabarcoding study of the mega-diverse Malay Archipelago.



5.2.5 Study sites

Figure 5.2. Lake sample sites across the Malay Archipelago. Each lake site is shown by red pins.

Study sites were selected to cover a range of biogeographical points across the Malay Archipelago, and to cover a range of geological lake formations, habitat variables and anthropogenic impact, details of which are shown below in Table 5.1. All lakes were sampled

from a boat, either via dugout canoe with no engine, small engine boats owned by local fishermen, or tourist style engine boats depending on the protection levels and infrastructure at each lake. The principal researcher (Alice Owusu-Evans) sampled the following sites with the help of members of the Indonesian Biodiversity Research Centre (IBRC): Lake (Danau) Batur, Lake (Danau) Beratan, Lake (Danau) Buyan, Lake (Danau) Semayang, Lake (Danau) Melintang, Lake (Danau) Rawa Pening and Lake (Danau) Matano. The following sites were sampled separately by members of the IBRC: Lake (Danau) Laut Tawar, Lake (Danau) Toba, Lake (Danau) Singkarak and Reservoir (Waduk) Riam Kanan. Lake (Tasik) Chenderoh was sampled by members of the Aquaculture Research Group at Universiti Sains Malaysia (University of Science Malaysia). Members of the IBRC and USM are fully credited in the Acknowledgements section.

Table 5.1. Lakes sampled in this study and associated data from the literature. (Arthana, 2011; GPS Coordinate Converter, 2018; Hardjamulia and Suwigno, 1988; Haryani, 2016; Kartamihardja, 2015; Kurniawan and Subehi, 2016; LakeNet, 2003a; LakeNet, 2003b; LakeNet, 2003c; LakeNet, 2003d; LakeNet, 2003e; LakeNet, 2003f; LakeNet, 2003g; LakeNet, 2003h; LakeNet, 2003i; LakeNet, 2003k; Lehmusluoto and Machbub, 1997; Mardiah and Syandri, 2016; Ministry of Environment Republic of Indonesia, 2012; Ministry of Environment Republic of Indonesia, 2012; Petr and Morris, 1995 Putri and Hadisusanto, 2016; Subehi *et al*. 2017; Tjahjo *et al*. 1998; Saragih and Sunito, 2001; UNEP, 2018; Whitten *et al*. 1996; Wijopriono *et al*. 2017.)

Lake	Lake Type	Location	Latitude	Longitude	Area (km^2)	Volume (km^3)	Max Depth (m)	Altitude (m)	Productivity
Batur	Enclosed, caldera	Kintamani, Bali, Indonesia	S 8° 15' 0" (-8.2500)	E 115° 24' 0" (115.4000)	16	0.82	88	1031	Mesotrophic-Eutrophic
Beratan	Enclosed, caldera	Tabanan, Bali, Indonesia	S 8° 16' 0" (-8.2667)	E 115° 10' 59" (115.1833)	3.85	0.049	22	1239	Mesotrophic-Eutrophic
Buyan	Enclosed, caldera	Buleleng, Bali, Indonesia	S 8° 14' 36.236" (-8.243399)	E 115° 7' 18.148" (115.121708)	3.9	0.116	87	1217	Mesotrophic-Eutrophic
Tamblingan	Enclosed, caldera	Buleleng, Bali, Indonesia	S 8° 15' 26.96" (-8.2574889)	E 115° 5' 46.852 (115.0963477)	1.9	0.027	90	1200	Oligotrophic
Matano	Tectonic	East Luwu, Sulawesi, Indonesia	S 2° 29' 29.431" (-2.4915087)	E 121° 22' 37.32" (121.3770336)	164	98	600	382	Ultraoligotrophic
Melintang	Floodplain, oxbow	East Kalimantan, Indonesia	S 0° 17' 37.537" (-0.2937602)	E 116° 20' 12.305" (116.3367514)	90	NA	5	10	Eutrophic
Semayang	Floodplain, oxbow	East Kalimantan, Indonesia	S 0° 18' 4.873" (-0.3013536)	E 116° 39' 19.206" (116.6553351)	240	NA	6.5	17	Eutrophic
Rawa Pening	Floodplain, semi-natural	Samarinda, Java, Indonesia	S 7° 17' 7.774" (-7.2854929)	E 110° 25' 55.801" (110.4321671)	25	0.052	14	470	Eutrophic
Singkarak	Tectonic	Solok and Tanah Datar, West-Sumatra, Indonesia	S 0° 37' 9.348" (-0.6192634)	E 100° 32' 27.103" (100.5408621)	107.8	16.1	268	360	Oligo-mesotrophic
Laut Tawar	Tectonic	Takengon city, Middle Aceh, Aceh, Indonesia	N 4° 36' 42.998" (4.6119439)	E 96° 55' 24.999" (96.9236109)	54.7	2.5	80	1200	Oligo-mesotrophic
Toba	Enclosed, caldera	North Tapanuli, Karo and Dairi regencies of Aceh, Indonesia	N 2° 47' 9.883" (2.7860786)	E 98° 36' 57.842" (98.6160674)	1,130	240	529	905	Oligotrophic
Riam Kanan	Reservoir	Tiwingan Lama, Aranio, Banjar, South Kalimantan, Indonesia	S 3° 31' 54.358" (3.531766)	E 115° 4' 3.054" (115.068201)	92	1.2	50	25	Mesotrophic
Tasik Chenderoh	Reservoir	Perak, Malaysia	N 4° 58' 18.788" (4.9718855)	E 100° 57' 34.226" (100.9595074)	8.5	0.095	16.2		Mesotrophic

5.2.6 Bali

Danau Batur

Figure 5.3. Lake Batur. Google Satellite image of Danau Batur, 2 km scale bar shown.

Lake Batur is the largest and deepest of the four Balinese lakes (see Table 5.1 for physical features), described as one of the world's largest and finest calderas (van Bemmelen, 1970). Lake Batur has markedly different physio-chemical features than the other Balinese lakes (see Table 5.1), with a much higher conductivity, and concentrations of magnesium, bicarbonate, chloride and sulphate, possibly due to the proximity of the active volcano Gunung Batur which most recently erupted in 1963 (Lehmusluoto and Machbub, 1997; Radiarta and Sagala, 2012; Sentosa and Wijaya, 2012; Haryani, 2016). The lake and its surrounding area is used by the local community for agriculture, tourism and fisheries, including the use of Floating Net Cages, particularly for tilapia (Lehmusluoto and Machbub, 1997; Arthana, 2011; Radiarta and Sagala, 2012), and is an important water storage source (Arthana, 2011). Uncontrolled land use change, particularly close to the lake's beach has caused high volumes of pollutants to enter Lake Batur (Arthana, 2011).

Non-native fish have been introduced to Lake Batur in an attempt to increase fishery activity in the region, the most common being Nile Tilapia (*Oreochromis niloticus*) (Sentosa and Wijaya, 2012). This species dominates fish catches with 63.96% of the catch according to a 2011 study (Sentosa and Wijaya, 2012), followed by Mozambique Tilapia (*Oreochromis mossambicus*) (13.63%) and the Yellow Rasbora (*Rasbora lateristriata*) (11.87%). However, production of Nile Tilapia is decreasing due to infection with some potentially pathogenic bacteria. The Grass carp (*Ctenopharyngodon idella*) was also introduced in 2009, a species generally introduced for controlling aquatic weeds such as the Common Water Hyacinth (Kartamihardja, 2012). In the same year, the Milk fish (*Chanos chanos*), (*Eichhornia crassipes*) (Kartamihardja, 2012) was introduced. Other fish species described from Lake Batur are shown in Appendix 7, although this is not exhaustive. Aquaculture activities in

Lake Batur are plenty, with Floating Net Cages (FNC) being implemented around the edge of the lake since the 1990s, growing to around 560 FNCs managed by 950 fish farmers (Suryaningtyas and Ulinuha, 2016). Higher nutrient concentrations were found closer to settlements and aquaculture cages, indicating their effect on potential eutrophication (Radiarta and Sagala, 2012). The use of organic and inorganic fertilizer has caused the nutrient content in Lake Batur to increase so that it is now classed as eutrophic, with the effect of nutrient, waste and pollutant influx exacerbated by the lack of an inlet or outlet (Arthana, 2011).

Danau Beratan (Danau Bratan)

Figure 5.4. Lake Beratan. Google Satellite image of Danau Beratan, 1 km scale bar shown.

Lake Beratan is the shallowest of the enclosed lakes of Bali (see Table 5.1), with a steep and rocky to the east where the caldera wall remains, and gently sloping and shallow to the west where there is a wave-cut platform (Green *et al*. 1978). Lake Beratan is under heavy pressure from recreational lake tourism related activities, including the use of high-power motorboats which cause engine oil pollution (Lehmusluoto and Machbub, 1997). There is a temple, Pura Ulun Danu (goddess of the lake) which offers a major tourist and religious attraction, and some small-scale agriculture nearby (Lehmusluoto and Machbub, 1997). The water is weakly stratified (RTR 19.0, 41.5 and 72.6), with a particularly low conductivity, and shows signs of eutrophication. The Convict Cichlid (also known as the Zebra Cichlid) (*Amatitlania nigrofascia*) is commonly found in Lake Beratan (Rahman *et al*., 2012; Sentosa *et al*. 2013; Restu *et al*. 2016), one of more than nine species of fish that were introduced since 1945 (Whitten *et al*. 1996), which are expected to be detrimental to the native fish community original and local fishing activities. Beratan Lake is known to contain the entirely endemic species *Rasbora baliensis* found only in this lake (Kottelat *et al*., 1993; Whitten *et al*. 1996) and so the presence of these introduced fishes and their impact on populations of *Rasbora baliensis* is of urgent concern (Whitten *et al*. 1996). In 1990 recreational fishermen noticed necrosis on the bodies of fish caused by a pathogenic bacteria, and by 1992, fish also carried *Lernaea* parasites, thought to be a result of the introduction of unhealthy fish stock and cumulative stress from pesticide loads (Whitten *et al*. 1996).

Danau Buyan

Lake Buyan is surrounded by rain forest, small-scale agriculture and quiet tourist accommodation, with low level fishing activities taking place (Lehmusluoto *et al*. 1997). The northern shore is formed by the steep, forested caldera wall, and the southern shore is gently sloping and cultivated, with a deepening basin to the west (Green *et al*. 1978). Lake Buyan is thought to have previously been connected to Lake Tamblingan, and separated after a land slide in 1818. Motorboats and water sports are banned on both lakes (Insight Guides, 2014). Fisheries development at Lake Buyan has been implemented to aid economic empowerment of the communities around the lake, nature tourism in the form of recreational fishing, and biodiversity conservation (Restu *et al*. 2016).

Figure 5.5. Lake Buyan. Google Satellite image of Danau Buyan, 1 km scale bar shown

A study from 2016 (Restu *et al*. 2016) found nine species of aquatic plants, and six species of fish (Appendix 4). The composition of the fish species found was *Amatitlania nigrofascia* (Convict Cichlid, 66%, introduced pest species), *Ostheocillus hasselti* (14%, the only native fish species), *Cyprinus carpio* (Common Carp 13%), *Oreochromis mosambicus* (Mozambique Tilapia 5%), and *Oreochromis niloticus* (Nile Tilapia, 0.3%). Nearby agricultural and tourist activities may have increased the nutrient load to this lake, contributing to the growth of aquatic plants (Restu *et al*. 2016). Organochlorine pesticide contaminants, i.e. DDT 5.02 ppb (parts per billion) and chlorotalonile 1.99 ppb were observed from 55 sampling point of water taken from five sampling zones, although neither were above maximum thresholds of 42 ppb (Manuaba, 2007).

5.2.7 Sulawesi

Danau Matano (Mantana)

Lake Matano (also known as Lake Mantana) is one of the ancient tectonic lakes found on the island of Sulawesi, the deepest lake in Southeast Asia, and eighth deepest lake in the world, at 590-600 m, with a cryptodepression of 218 m (see Table 5.1). It is the hydrological head of the Matano-Mahalona-Towuti chain in the Malili Lakes system, thought to be 41-12 million years old (Brooks, 1950; Whitten *et al*. 1987; Haffner *et al*. 2001; Nasution, 2016). It flows into nearby Danau Mahalona, which in turn flows into Danau Towuti, before emptying into the Gulf of Bone in east Sulawesi (LakeNet, 2003d; Herder *et al*. 2012). Lake Matano has a sharp thermocline layer, and at 150-200 m deep, a clear physical and chemical gradient where alkalinity and calcium increases two-fold, magnesium three-fold, iron and total nitrogen ten-fold, ammonia twenty-fold, and manganese from undetected to 0.22 mg/l, while sodium decreases from about 3 to 1.1 mg/l (Lehmusluoto and Machbub, 1997).

Figure 5.6. Lake Matano. Google Satellite image of Danau Matano, 2 km scale bar shown

There is an anoxic hypolimnion, with weak stratification (RTR 34.3) and 7.4 x 105 metric ton of CH4 (Lehmusluoto and Machbub, 1997; Crowe *et al*. 2010). It provides a water source, tourism attraction, fishing opportunities and is designated a World Heritage site and National Tourism Park (Nasution, 2006), as well as a LakeNet Biodiversity Priority and WWF Global 200 ecoregion (LakeNet, 2003d). These ancient lakes harbour endemic radiations of a variety

of freshwater taxa, including fishes, molluscs, shrimps and crabs (von Rintelen *et al*. 2012), which have provided model systems from which to explore the adaptive character of intralacustrine radiations (von Rintelen *et al*. 2004; Herder *et al*. 2006, 2008; Pfaender *et al*. 2010, 2011), behavioural specialization and filial cannibalism (Gray *et al*. 2007, 2008a; Cerwenka *et al*. 2012) and male colour polymorphisms (Gray *et al*. 2008b; Walter *et al*. 2009) amongst other evolutionary topics. As an ultraoligotrophic lake with very low productivity, its waters are crystal clear (Crowe *et al*. 2008). Fish species found there include *Telmatherina sarasinorum* (Nilawati *et al*. 2010), flowerhorns (*Amphilophus* sp., Herder *et al*. 2012). There are fourteen endemic fish species, *Telmatherina antoniae* dominates the fish population at Lake Matano, followed *Glossogobius matanensis* (Wirjoatmodjo *et al*. 2003; Nasution, 2016). The high degree of endemism in the fish communities of Lake Matano and its neighbours justifies the need for freshwater biodiversity conservation. Threats include a hydroelectric power plant, ornamental fish trade, fishing, ecotourism, introduction of invasive species, habitat degradation through soil erosion, logging, mining and agriculture, and transportation, particularly from the nearby nickel industrial plant owned by PT Inco. PT Inco is running the largest nickel laterite ore operation in the world in Sulawesi, with Lake Matano being one of its sites (Nasution, 2006; Haryani, 2016). Deforestation poses a major threat to the Malili Lake system, aggravated by government directed population relocation from greater Sunda Islands to less densely populated areas of Sulawesi, possibly leading to increased run-off, higher nutrient pollution and consequent eutrophication (LakeNet, 2003d).
Kalimantan

5.2.8 East Kalimantan, Borneo

Danau Melintang and Danau Semayang

Lake Melintang and Semayang are two of the cascading, floodplain, oxbow, eutrophic lakes (Petr and Morris, 1995) connected to the Mahakam River along with Jempang, in East Kalimantan. These lakes are shallow with a muddy, sandy floor (see Table 5.1), with a fish community dominated by the Cyprinidae family (LakeNet, 2003e; Haryani, 2016; Kurniawan and Subehi, 2016). The lakes connected by the Mahakam River have been particularly affected by heavy metal pollutants, silting, and river-borne erosion, causing habitat loss, disruption of reproductive processes in aquatic animals and the growth of water hyacinth (LakeNet, 2003g; Kartamihardja, 2015; Hiryani, 2016; Kurniawan and Subehi, 2016).). Around 75% of East Kalimantan has been assigned for coal mining (Green Peace, 2016), the activity of which can be observed near Danau Melintang and Danau Semayang sampled in

this study, through satellite imagery using Zoom Earth (2018), Google Earth (2018) and Google Street View (2018) (Figure 5.7 below). The mining company 'PT. Gema Rahmi Persada', (Desa Kotbangun2, Kotabangun, kutai kartanegara, East Borneo), operate the mine site at nearby Kotabangun (Four Square, 2018) which appears to be responsible for the visible red pollution entering the lake system, likely through acid mine drainage, which creates contaminants in the form of acid, iron, sulphur and aluminium, which can cause loss of aquatic life, and restricts stream use for recreation, public drinking water and industrial water supplies (U.S. Environmental Protection Agency, 2018). Bright blue water in abandoned open-pit mines are visible, likely an indicator of highly acidic waste water (Green Peace, 2016). Indeed, there are reports of local people living on the nearby Santan River abandoning their homes because of the level of degradation of the river and water quality, which has deteriorated to the level that it is necessary for local people to buy bottled water.

Figure 5.7. Lake Semayang and Melintang. Google Satellite image of Danau Semayang and Melintang, 5 km scale bar shown.

Responsibility is shirked by the local mining company, PT Indominco Mandiri, (owned by the larger Thai mining company, Banpu) who's CEO has stated that mining activities complied with environmental regulations (Mongobay, 2017; Tisnadibrata and Wiriyapong, 2016). A report from Greenpeace (2016) also describes intense mining impacts from PT Mahakam Sumber Jaya (MSJ, Harum Energi Group), in other villages near the Mahakam River, and the imprisonment of local farmers who peacefully protested against them.

Unsurprisingly, local populations of the critically endangered Irrawady River Dolphin (*Ocaella brevirostris*) have decreased (Haryani, 2016).

Waduk Riam Kanan

Figure 5.8. Google Satellite image of Waduk Riam Kanan. 2 km scale bar shown.

The Banjar Regency in South Kalimantan is dominated by rivers, with the capital city, Banjarmasin, known locally as Kota Seribu Sungai (Indonesian: City of Thousand Rivers). Many people live around the Martapura, Barito, and Riam Kanan Rivers. Freshwater fisheries are therefore an important source of food locally, with high demand for fish, resulting in a high level of local aquaculture activity (Hidayaturrahmah, 2017). The Riam Kanan reservoir was constructed in 1973 (Kartamihardja, 2015) by damming the Riam Kanan River (MacKinnon, 1996) to act as a field, irrigation and electricity reservoir (Hardjamulia and Suwigno, 1988). Floating-net culture began in 1986 (Hardjamulia and Suwigno, 1988; Rahman *et al*. 2017) and nearby villages grow rice, peanuts and vegetables in small plots at the forest edge, although it is nearby mining activities which contribute to sediment influx and pollution (MacKinnon, 1996). In 2001, the Riam Kanan River hosted around 6,800 fishcages, which decreased to 4,667 in 2006 as a result of fish mortality related to over intensification (Rahman *et al*. 2017).

5.2.9 Java

Danau Rawa Pening

Figure 5.9. Lake Rawa Pening. Google Satellite image of Danau Rawa Pening, 2 km scale bar shown

Lake Rawa Pening is a shallow, flood-plain, semi-natural, eutrophic lake in Central Java. Lying within an ancient caldera, Lake Rawa Pening is a man-made lake, created by a control dam on the Tutang river between 1921 and 1923, heavily exploited for fisheries and other water related economic activities (bottom mud and molluscs) and surrounded by large areas of rice paddies and towns (Lehmusluoto and Machbub, 1997). There is no current management plan, and no acknowledgement by any government body of responsibility for management of the lake (UNEP, 2018). However, Whitten *et al*. (1996) state that no area of freshwater in Indonesia is better studied ecologically than Rawa Pening, with many early studies of its physical, chemical, biological and sociological features. It is Indonesia's oldest reservoir, with an inlet through the Muncul estuary where many fish go to spawn (Whitten *et al*. 1996), and an outlet to the Tuntang River (Lehmusluoto and Machbub, 1997). Invasive Water Hyacinth (*Eichornia crassipes*, known locally as enceng gondok) infects 40-60% of the lake surface (Hutarabat *et al*, 1986; Lehmusluoto and Machbub, 1997; UNEP, 2018), introduced as green manure into nearby rice fields (Whitten *et al*. 1996). The introduced *Anodonta woodiana* can be found at Lake Beratan, where it is consumed for food (Whitten *et al*. 1996). Fish yields dropped from 548 to 18 kg/ha between 1972 and 1980, likely due to overfishing, but have recovered since 1980, possibly due to floating cages (Whitten *et al*. 1996). The lake is fed by nine rivers running down nearby slopes, and by a number of internal springs (Irawan, 2016). Large amounts of allochthonous matter from the catchment come from the nearby towns of Salatiga and Ambarawa which increase run-off of untreated plastic and organic waste which pollute the lake inlets (Lehmusluoto and Machbub, 1997; UNEP, 2018), as well as clogging from water hyacinth (*Eichhornia crassipes*) (Irawan, 2016). This lake also provides hydroelectric power, irrigation, recreational services, drinking

water and fishing activities (Irawan, 2016). The lake has no epilimnion, but a thermocline which begins at the surface (Lehmusluoto and Machbub, 1997; UNEP, 2018).

5.2.10 Sumatra

Danau Singkarak

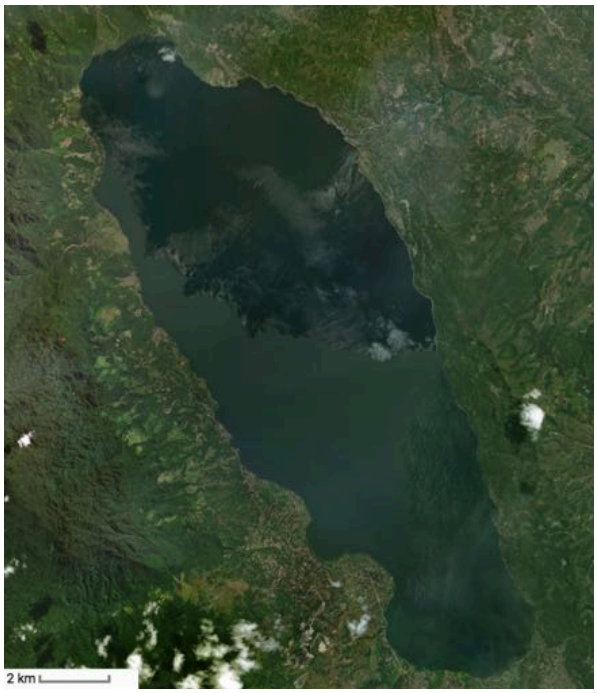


Figure 5.10. Lake Singkarak. Google Satellite image of Danau Singkarak, 2 km scale bar shown

Lake Singkarak is a strike slip fault, tectonic, oligotrophic lake (Petr and Morris, 1995) located in Solok and Tanah Datar, West-Sumatra, with a natural flushing system (inlet from Dibawah lake via River Sumani/outlet through Umbilin river), and functions as a sediment sink (Lehmusluoto and Machbub, 1997). Lake Singkarak is a popular tourist lake, part of an annual international tourist event (Oktavia and Faoziyah, 2016). River inlets come from the Sumpur River, Paninggahan River, Raing River, Muara Pingai River, Saning Bakar River and Sumani River (Syandri, 1996), and since 1998 the outlet flows through a hydropower tunnel (Mardiah and Syandri, 2016). Studies from 2013 and 2016 found 19 fish species, belonging to 5 orders, 9 families, 16 genera (Syandri *et al*. 2013; Oktavia and Faoziyah, 2016) (see Appendix 7). The lift net survey conducted by Mardiah and Syandri (2016) yielded a total catch constituting of: Cyprinidae (42.10%), Bagridae (10.52%), Osphronemidae (10.52%), Channidae (10.52%), Tetrodontidae (5.26%), Anabantidae (5.26%), Mastacembelidae (5.26%), Chiclidae (5.26%), and Gobiidae (5.26%). These were composed of bilih fish (*Mystacoleucus padangensis*) (81.17%), the Tinfoil Barb (*Barbonymus schwanenfeldii*) (4.26%), the Hampala Barb (*Hampala macrolepidota*) (5.34%), a barb species, (*Anematichthys armatus*) (1.70%), a crustacean species within the *Penaeus* genus (6.68%), and the Humpback Puffer (*Tetraodon palembangensis*) (0.70%). Lake Singkarak shows stratification (RTR 55.2 and 78.8), with a permanently or semi permanently stagnant hypolimnion, and is likely meromictic from around 45- 50 m, meaning around two thirds of the lake is oxygen depleted (Lehmusluoto and Machbub, 1997). Threats

to this lake include plans for abstraction for hydroelectric dam development and irrigation, potentially causing mixing of the stratified layers (Lehmusluoto and Machbub, 1997).

Danau Laut Tawar

Figure 5.11. Lake Laut Tawar. Google Satellite image of Danau Laut Tawar 2 km scale bar shown.

Lake Laut Tawar is a large, tropical, subalpine, eutrophic lake located in the eastern area of Takengon city, Middle Aceh, Aceh, Indonesia (LakeNet, 2003j; Lumbantobing, 2010; Putri and Hadisusanto, 2016). The lake is a water source and fishing grounds for fishermen of the Gayoness people. There are at least 25 short tributaries flowing into the lake, and only one outlet through the Peusangan River. The lake is surrounded by almost barren pine forest, and mountains which reach above 200 m, and lies on a substrate of granite rock (LakeNet, 2003j; Putri and Hadisusanto, 2016). This lake has unique environmental conditions, characterized by high light intensity throughout the year, low air temperature, high rainfall and strong winds (Putri and Hadisusanto, 2016). Floating cage culture activity is found in high amounts, increasing nutrient loads and decreasing transparency. There is stratification, with an epilimnion 0 – 5m, metalimnion 5 – 8m, and hypolimnion less than 15m. Threats include illegal logging, tourism, global warming and other human activities, which have resulted in decreased water quality and quantity, possibly adversely affecting fishes. The watershed is covered by forests, which are increasingly affected by deforestation, and agricultural activities. (Putri and Hadisusanto, 2016). There is a high level of endemism of freshwater fishes in North-western Sumatra compared to other regions in Sundaland (Roberts, 1989; Kottelat, 1994), including four new species of *Rasbora*, including the aptly named *Rasbora tawarensis* (see Appendix 7) (Lumbantobing, 2010).

Danau Toba

Danau Toba, spread across the North Tapanuli, Simalungun, Karo and Dairi regencies of Aceh, Sumatra is the largest natural lake in Indonesia, and the largest volcanic lake in the world (LakeNet, 2003k; Kurniawan and Subehi, 2016). It is a volcano-tectonic, oligotrophic lake (Petr and Morris, 1995) formed as a caldera eruption from Mount Toba around 75,000 years ago, (Ninkovich *et al*. 1978), leaving the 'pseudo-island' of Samosir island in the centre, surrounded by hills and mountains up to 2000 m above sea level (LakeNet, 2003k; Haryani, 2016; Pratiwi *et al*. 2016). There are 202 inlets to the lake, 70 of which run year-round (LakeNet, 2003k), and one outlet through the River Asahan at Porsea to the Strait of Malacca (Lehmusluoto and Machbub, 1997). The northern basin may experience periodic circulation (RTR 31.1 and 61.1), but the southern basin has a clearer thermocline and oxycline at 100-150 m, and so is more likely to remain stagnant throughout the year (RTR 25.0), with stratification at 140 – 50 m, and oxygen depletion at 0.061 mg/l (Lehmusluoto and Machbub, 1997). There are 27 fish species listed from Danau Toba on Fishbase (2017a), shown in Appendix 7. There are two endemic fish found in Danau Toba, *Rasbora tobana* (Lumbantobing, 2010) and *Neolissochilus thienemanni* (Saragih and Sunito, 2001). Fish surveys in 1990 showed a species composition dominated by Cyprinidae (*Barbodes gonionotus, Cyprinus carpio, Mystacoleucus padangensis, Barbodes binotatus* and *Rasbora jacobsoni*) (Wetlands International Indonesia, 1990) with a family distribution of 31.25% Cyprinidae, 12.5% Clariidae, 12.5% Cichlidae, 12.5% Channidae, 12.5% Belontiidae, 6.25% Aplocheilidae, 6.25% Poeciliidae and 6.25% Osphronemidae (Kurniawan and Subehi, 2016).

Figure 5.12. Lake Toba. Google Satellite image of Danau Toba 10 km scale bar shown.

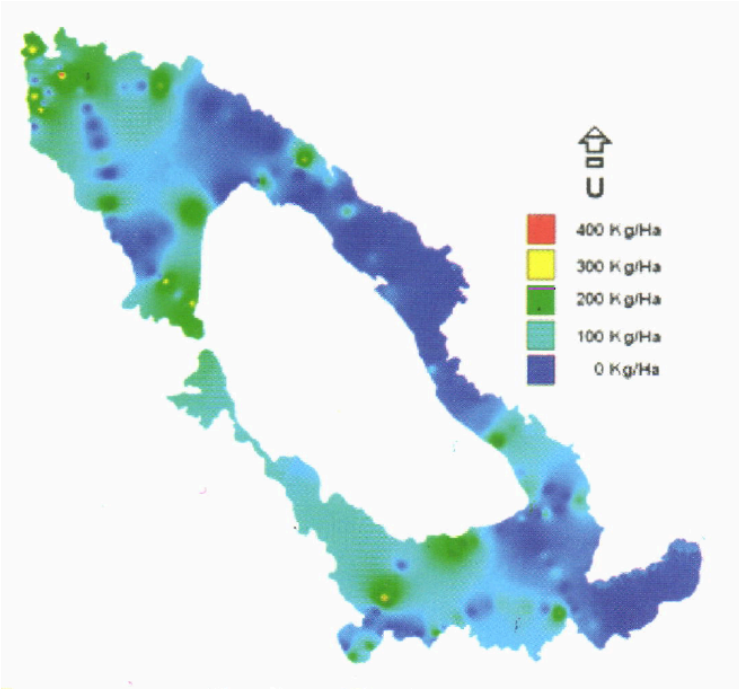


Figure 5.13 Variation of fish abundance in Lake Toba. October 2005 (Wijopriono *et al*. 2017)

The Common Carp was introduced from 1905, along with the Mossambique Tilapia (*Oreochromis mossambicus*) in the 1940s, and the Nile Tilapia (*Oreochromis niloticus*) by the 1950s (Kartamihardja, 2012). The Giant Gourami (*Osphronemus goramy*) and Snake Skin Gourami (*Trichogaster pectoralis*) were also introduced in the 1920s, although these did not become established (Kartamihardja, 2012). In 2003, around 3,000 heads of 'bilih' (*Mystacoleucus padangensis*) were introduced to Lake Toba from nearby Lake Singkarak to increase the lake's fish production, which became dominated by these fish (Kartamihardja, 2012; Hedianto and Kartamihardja, 2014; Kartamihardja, 2015; Kartamihardja *et al*. 2015). In 2013 however, an interesting phenomenon occurred in which populations of *M. padangensis* sharply decreased, followed by an increase of the introduced, and economically unprofitable Glassfish (*Parambassis siamensis*) which preys upon the eggs of *M. padangensis*. Larger fish were found in the Northern and Western areas of Samosir Island, and more often in deep water, and the Eastern and Southern areas, and shallow waters contained smaller fish. There was a higher biomass of fish in the Northwestern and Southwestern area of the lake, as shown by the heatmap in Figure 5.13 above (Wijopriono *et al*. 2017).

5.2.11 Peninsular Malaysia

Tasik Chenderoh

Tasik Chenderoh is a reservoir lake located on the Perak River in the state of Perak, Peninsular Malaysia. It is the oldest reservoir in Malaysia, created for hydroelectric power in 1930 (Dahlen 1993). It is a mesotrophic reservoir, with impacts on fish communities from water level management and fluctuation, riparian land development, and housing

Figure 5.14. Google Satellite image of Tasik Chenderoh. 2 km scale bar shown.

developments (Ali, 1996). A change in fish community composition has occurred after the impoundment of the lotic ecosystem and conversion to a lentic ecosystem, and the consequent anthropogenic effects including water level regulation, shoreline development and the installation of cage culture (Ali, 1996). Fish species from Tasik Chenderoh are shown in Appendix 7. The highest catch among commercial species was for *Puntioplites bulu* (14.1%), Mystus sp. (10.0%), *Thynnichthys thynnoides* (8.8%) and Channidae sp. (5.7%), and for non-commercial species > 20% of the total catch were from *Barbonymus schwanenfeldii*, > 15% from *Cyclocheilichthys apogon*, and > 10% *Osteochilus vittatus* (Kah-Wai and Ali, 2000). Perak is one of the few states in Peninsula Malaysia which has implemented Inland Fishery Regulations limiting the types of gear used in the fishing activity, controlling the use of destructive fishing techniques including poisoning, electro-fishing and small mesh gill nets as of the 1980s (Kah-Wai and Ali, 2000).

5.2.12 Aims and Objectives

To our knowledge, this is the first study to use eDNA metabarcoding of lake water samples from Southeast Asia. The broad aim of this study is to assess the potential of this low-effort approach for multi-species fish detection and wider biodiversity of the lakes of the Malay Archipelago, and examine whether patterns of species community composition and richness vary with habitat variables in these freshwater systems. It is expected that larger lakes host more species and areas East of the Wallace line have a distinctly unique community composition, including an absence of cyprinid fish. We used previously published primer sets targeting the COI, 12S, and 16S regions of the mitochondrial genome. This study represents the first targeted effort that demonstrates the effectiveness of an eDNA metabarcoding approach for the detection and monitoring of Southeast Asian fish communities and aquatic biodiversity.

Aims:

a) To assess the ability of eDNA to monitor aquatic biodiversity from each lake across the Malay Archipelago based on OTU clusters amplified from eDNA samples.

b) To investigate how species richness and composition relate to lake habitat variables including altitude, lake size, lake depth, productivity and region.

c) To characterise how OTU richness, composition and species assignment relate to biogeography, as eDNA information should reflect the presence of local species.

5.3 Methods

This study used eDNA metabarcoding to assess freshwater biodiversity. This is a multi-species approach, by combining eDNA sampling with universal multi-gene metabarcoding, so that broad biodiversity information can be generated without *a priori* information, although amplification bias and primer specificity can limit reliability. This approach is in contrast to single-species methods in which one, or several, species of interest are targeted using species-specific primers, which yields less information and requires a priori information about the target species, but may be more specific in terms of DNA amplification. The eDNA sampling, laboratory and bioinformatic methods employed are described in the Universal Methods (Chapter 4). For this study, the sampling effort was designed to increase with increasing lake size, to allow for the levels of heterogeneity within a single lake environment. A 750 ml sub-sample was collected every 500 m for 2.5 km using

a plastic jug, and combined into a plastic bucket (both the jug and bucket were sterilised using 20% bleach and rinsed with 70% ethanol prior to sampling). Each 2.5 km transect therefore consisted of 6 x sub-samples from 0, 500, 1,000, 1,500, 2,000, and 2,500 m combined into one large sample. From this one large sample, 3 x Sterivex filter replicates were used to filter 500 ml each, totalling 1.5 L filtered from the 4.5 L collected. The rationale for these 6 x subsamples being combined into one large sample was to cover the maximum area for eDNA collection for the number of Sterivex filters available. The smallest lakes sampled in this way (Danau Beratan and Danau Buyan) were only 2.5 km long, and so only one transect (3 x Sterivex filters) was completed. Because of this, Danau Beratan was sampled on three occasions a few days apart to check for consistency in eDNA results. The largest lake, Danau Toba, is 100 km long, and so it was not possible to sample the entire length of the lake, but instead 5 x transects were sampled in the North of the lake, and 5 x transects sampled in the South of the lake, totalling 25 km (10 x 2.5 km transects) and 30 x Sterivex filters. The number of transects which yielded acceptable metabarcoding information are shown in Figure 5.16, 5.17 and 5.18 indicated by the lake name followed by a number.

5.3.1 Statistical methods

An Analysis of Variance (ANOVA) was used to compare OTU richness between sites. Non-Metric Multidimensional Scaling (NMDS) plots were created using vegan (Oksanen *et al*. 2013) and ggplot (Wickham, 2016) in R using the Total Marker data category. NMDS plots were created in combination with the Manhattan dissimilarity method for calculating a distance matrix (chosen by vegan based on the dataframe of these OTUs). Several variables were separately incorporated, consisting of Region, Area, Max Depth, Productivity and Altitude. 'Region' was the geographic region of the lakes where the sample was collected, either Bali (lakes Beratan, Batur, and Buyan), Sumatra (lakes Toba and Laut Tawar), Java (lake Rawa Pening), Sulawesi (Lake Matano), Kalimantan (lakes Melintang, Semayang and Riam Kanan) or Malaysia (the Chenderoh Reservoir). The 'Area' variable grouped lakes into 'large, 'medium or 'small based on the area described from the literature, and the OTU richness compared. Large lakes were those with an area of more than 100 km^2 (n = 3 lakes, 18 transects), medium lakes were those with an area 11 – 100 km^2 (n = 4 lakes, 19 transects), and small lakes were those with a maximum depth between 0 – 10 km^2 (n = 4 lakes, 10 transects). 'Max Depth' was the maximum recorded depth of the lake, either 'deep' with maximum depth of more than 100 m (n = 2 lakes, 14 transects), 'medium' with a maximum

depth of 21 – 100 m (n = 5 lakes, 19 transects), or 'shallow' with a maximum depth between 0 – 20 m (n = 4 lakes, 14 transects). 'Productivity' was the trophic state of the lake identified from the literature, either 'eutrophic', 'mesotrophic' or 'oligotrophic'. 'Altitude' was the height above sea level at which the lake resided, either 'highland' (altitude of > 1000 m above sea level, n = 4 lakes, 13 transects), 'midland' (altitude of > 101 - 999 m above sea level, n = 3 lakes, 17 transects) or 'lowland' (altitude of 0 – 100 m above sea level, n = 4 lakes, 17 transects). A Permutational Analysis of Variance (PermANOVA) was also performed on the NMDS OTU distance tables using ADONIS from the vegan package in R with 999 permutations.

5.4 Results

Table 5.2. Summary of OTU richness.

Marker	**All**	**COI**	**12S**	**16S**
Maximum OTU Richness	76	64	27	19
Average OTU Richness	39	31	6	3
Minimum OTU Richness	13	5	0	0

The highest OTU richness from all markers combined was 76 OTUs from a transect from Lake Semayang, and the lowest was 13 from a transect from the Riam Kanan reservoir. The highest and lowest OTU richness from the COI data was 64 from a transect from Lake Semayang, and 5 from a transect from Rawa Pening respectively.

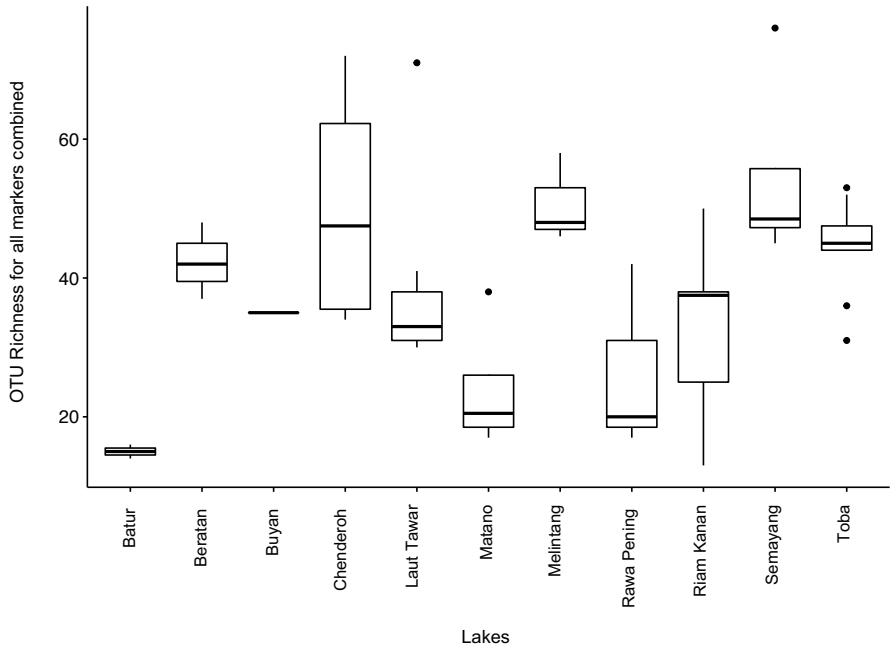


Figure 5.15. OTU richness per lake. OTU richness for all markers combined per lake.

OTU richness was compared between groups using an Analysis of Variance (ANOVA) in R. For all markers combined, there was a significant difference between lakes (P = 0.002).

Table 5.3 OTU richness per lake. Average, minimum and maximum OTU richness per lake sorted in descending order per category. Consistently low OTU richness was observed in Lakes Batur, Matano and Riam Kanan, and consistently high OTU richness in Lakes Semayang, Chenderoh and Melintang.

Measure	Lake	All OTUs	Lake	12S OTUs	Lake	16S OTUs	Lake	COI OTUs
Average	Batur	15	Riam Kanan	1	Batur	0	Batur	12
Average	Matano	24	Toba	2	Matano	0	Matano	20
Average	Rawa Pening	26	Batur	3	Rawa Pening	0	Rawa Pening	21
Average	Riam Kanan	33	Matano	4	Toba	1	Laut Tawar	26
Average	Laut Tawar	39	Rawa Pening	5	Riam Kanan	1	Chenderoh	27
Average	Beratan	42	Laut Tawar	8	Semayang	2	Melintang	31
Average	Toba	44	Beratan	8	Beratan	2	Riam Kanan	31
Average	Chenderoh	50	Semayang	10	Laut Tawar	5	Beratan	32
Average	Melintang	51	Melintang	13	Melintang	7	Toba	42
Average	Semayang	55	Chenderoh	15	Chenderoh	9	Semayang	43
Max	Batur	16	Batur	3	Batur	0	Batur	13
Max	Matano	38	Riam Kanan	5	Matano	0	Laut Tawar	30
Max	Rawa Pening	42	Toba	6	Rawa Pening	1	Matano	31
Max	Beratan	48	Matano	7	Toba	2	Beratan	36
Max	Riam Kanan	50	Beratan	10	Beratan	3	Melintang	36
Max	Toba	53	Rawa Pening	11	Riam Kanan	3	Chenderoh	38
Max	Melintang	58	Semayang	12	Semayang	5	Rawa Pening	39
Max	Laut Tawar	71	Melintang	14	Melintang	8	Riam Kanan	42
Max	Chenderoh	72	Chenderoh	23	Chenderoh	16	Toba	52
Max	Semayang	76	Laut Tawar	27	Laut Tawar	19	Semayang	64
Min	Batur	14	Matano	0	Batur	0	Rawa Pening	5
Min	Riam Kanan	13	Riam Kanan	0	Toba	0	Batur	11
Min	Matano	17	Toba	1	Matano	0	Matano	13
Min	Rawa Pening	17	Rawa Pening	1	Semayang	0	Riam Kanan	13
Min	Laut Tawar	30	Batur	3	Riam Kanan	0	Chenderoh	19
Min	Toba	31	Laut Tawar	3	Rawa Pening	0	Laut Tawar	23
Min	Chenderoh	34	Beratan	5	Beratan	2	Melintang	28
Min	Beratan	37	Chenderoh	6	Laut Tawar	2	Toba	29
Min	Semayang	45	Semayang	7	Chenderoh	4	Semayang	29
Min	Melintang	46	Melintang	12	Melintang	6	Beratan	30

Lake Singkarak was removed from the analysis as samples were degraded due to an error with the shipping provider from Indonesia to Europe. In the 12S dataset, only one OTU remained from Lake Singkarak after filtering, with 936 raw reads assigned to OTU 100, which returned 'no hits' from the MEGAN assignment pipeline and 'No significant similarity found' using a BLAST with a 'highly similar sequence' search (megablast), and a range of distantly related species with low quality BLAST criteria when using a search of 'somewhat similar' sequences (blastn)'.

5.4.1 12S Marker

There were 87 high quality OTUs produced after bioinformatic filtering shown below in Table 5.4 of which 29 could be assigned to species level, and 40 could be assigned to genus level, from 14 orders. Of the 14 orders, 13 were from the ray-finned fish class Actinopterygii, and one from the class Mammalia. Taxonomic orders consisted of Anabantiformes (11), Carangiformes (2), Characiformes (1), Cichliformes (12), Clupeiformes (2), Cypriniformes (33), Cyprinodontiformes (4), Perciformes (7), Gonorynchiformes (1), Osteoglossiformes (2), Scombriformes (1), Siluriformes (8), Synbranchiformes (1), Primates (1). The order with the highest number of OTUs was Cypriniformes (33), one of the most abundant fish groups in Southeast Asia. A transect from the Chenderoh Reservoir in Malaysia had the highest OTU richness of 27, whilst a transect from Lake Matano and several from Riam Kanan had the lowest of 0. Many of the fish species detected, both native and introduced, are economically significant aquaculture or fisheries species. The Middle Eastern and African cichlid fishes belonging to the Pseudocrenilabrinae subfamily, including the *Oreochromis* and Sarotherodon genera were common across all lakes except for Semayang and Melintang in Borneo, and particularly dominant in the Balinese lakes and Rawa Pening in Java (see Figure 5.16 below). The Common Carp (*Cyprinus carpio*) was also common, found in five lakes, as was an OTU assigned to the *Osteochilus* genus, likely to be *Osteochilus vittatus*.

Lake Toba was almost entirely dominated by an OTU assigned to the Ambassidae family (see Figure 5.16 below), very likely to be the introduced Glassfish *Parambassis siamensis.* Similarly, the Chenderoh Reservoir was mostly dominated by the Perak River Sprat (*Clupeichthys perakensis*), although at much lower read counts, with many more coexisting species (Figure 5.16).

Reads from lakes Melintang and Semayang in Borneo were mostly dominated by an OTU assigned to the *Helostoma* genus (most likely *Helostoma temminckii*).

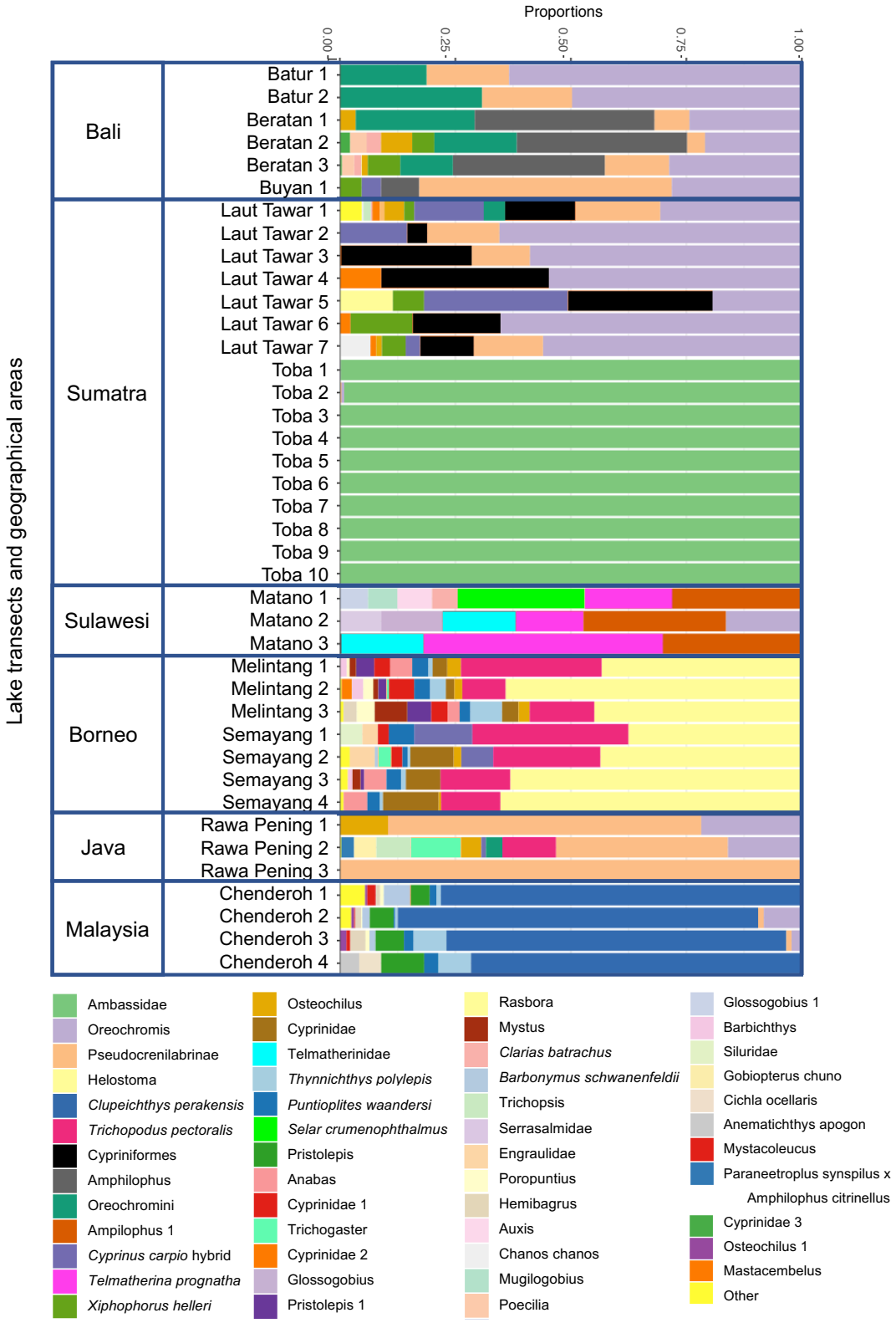


Figure 5.16. Taxa plot of 12S marker reads across all lakes. This plot was created using the top 50 most abundant taxa selected from the species level and higher depending on the filtering criteria. Each lake shows a similar community composition, and it is clear that lakes in the same region (e.g. the lakes of Bali and the lakes of Borneo) show a similar community composition. The reservoir Riam Kanan was removed as it appeared blank in this taxa plot due to a lack of 12S reads.

Table 5.4. Taxonomic assignments of 12S OTUs, and their presence per lake. The number of transects with a positive detection are shown in each coloured cell. Colour coding is based on the total reads found from all samples from one lake after bioinformatic filtering but before normalisation. Red = > 10,000 reads across a single, orange = 5,000 – 9,999, yellow = 2,500 – 4,999, green = 1,000 – 2,499 and blue = 20 – 999. BA = Lake Batur, BE = Lake Beratan, BU = Lake Buyan, TO = Lake Toba, RP = Lake Rawa Pening, SE = Lake Semayang, ME = Lake Melintang, RK = the Riam Kanan Reservoir, MA = Lake Matano, CH = the Chenderoh Reservoir. Species with a * were accepted beneath the 90 - 98% identity threshold as they are the only species within the genus occurring in this geographic region.

	12S													
Order	**Family**	**Genus / Species**	**Common Name**	**BA**	**BE**	**BU**	**LT**	**TO**	**RP**	**SE**	**ME**	**RK**	**MA**	**CH**
Anabantiformes	Anabantidae	*Anabas testudineus**	Climbing Perch							2/4	2/3			
Anabantiformes	Channidae	*Channa micropeltes*	Giant Snakehead							1/4				
Anabantiformes	Helostomatidae	*Helostoma temminckii**	Gourami							4/4	3/3			
Anabantiformes	Osphronemidae	*Osphronemus*	Gourami									1/6		1/4
Anabantiformes	Osphronemidae	*Trichopodus pectoralis*	Snakeskin Gourami						1/3	4/4	3/3			
Anabantiformes	Osphronemidae	*Trichopodus*	Gourami				1/7		1/3	1/4	1/3			
Anabantiformes	Osphronemidae	*Trichopsis*	Gourami				1/7		1/3					
Anabantiformes	Pristolepididae	*Pristolepis*	Leaffish											4/4
Anabantiformes	Pristolepididae	*Pristolepis*	Leaffish							1/3	3/3			
Anabantiformes	Telmatherinidae	*Telmatherina prognatha*	Sail-fin silverside										3/4	
Anabantiformes	Telmatherinidae	-	Sail-fin silverside										2/4	
Carangiformes	Carangidae	*Decapterus macarellus*	Mackerel Scad				1/7							
Carangiformes	Carangidae	*Selar crumenophthalmus*	Big Scad										1/4	
Characiformes	Serrasalmidae	-	Characiform										1/4	
Cichliformes	Cichlidae	*Amphilophus*	Cichlid		2/3	1/1								
Cichliformes	Cichlidae	*Amphilophus*	Cichlid										3/4	
Cichliformes	Cichlidae	*Cichla ocellaris*	Peacock Bass											1/4
Cichliformes	Cichlidae	*Paraneetroplus synspilus x Amphilophus citrinellus*	Cichlid hybrid						1/3					
Cichliformes	Cichlidae: Tilapiini	-	Tilapia	2/2	3/3		1/7		1/3					

				BA	BE	BU	LT	TO	RP	SE	ME	RK	MA	CH
Cichliformes	Cichlidae	*Sarotherodon*	Tilapia				1/7	1/10						
Cichliformes	Cichlidae	*Oreochromis*	Tilapia				1/7							
Cichliformes	Cichlidae	*Oreochromis*	Tilapia	2/2	3/3	1/1	7/7	2/10	2/3			2/6	1/4	2/4
Cichliformes	Cichlidae	*Oreochromis*	Tilapia				1/7	1/10						
Cichliformes	Cichlidae	*Pseudocrenilabrinae*	Tilapia				1/7							
Cichliformes	Cichlidae	*Pseudocrenilabrinae*	Tilapia				1/7							
Cichliformes	Cichlidae	*Pseudocrenilabrinae*	Tilapia	2/2	3/3	1/1	4/7	1/10	3/3			2/6		2/4
Clupeiformes	Clupeidae	*Clupeichthys perakensis*	Perak River Sprat											4/4
Clupeiformes	Engraulidae	-	Anchovy							2/3				
Cypriniformes	Cyprinidae	*Cyclocheilichthys apogon*	Beardless Barb											2/4
Cypriniformes	Cyprinidae	*Barbichthys laevis*	Sucker Barb							1/4	2/4			
Cypriniformes	Cyprinidae	*Barbonymus*	Barb fish				1/7							
Cypriniformes	Cyprinidae	*Barbonymus gonionotus*	Silver Barb											1/4
Cypriniformes	Cyprinidae	*Crossocheilus*	Algae eater											1/4
Cypriniformes	Cyprinidae	*Barbonymus schwanenfeldii*	Tinfoil Barb							1/3				3/4
Cypriniformes	Cyprinidae	-	Cyprinid fish											1/4
Cypriniformes	Cyprinidae	-	Cyprinid fish				1/7							
Cypriniformes	Cyprinidae	-	Cyprinid fish				1/7							
Cypriniformes	Cyprinidae	-	Cyprinid fish				1/7							
Cypriniformes	Cyprinidae	-	Cyprinid fish							3/4	3/3			
Cypriniformes	Cyprinidae	-	Cyprinid fish							2/4	3/3			
Cypriniformes	Cyprinidae	-	Cyprinid fish								3/3			
Cypriniformes	Cyprinidae	-	Cyprinid fish				4/7							1/4
Cypriniformes	Cyprinidae	-	Cyprinid fish		2/3									
Cypriniformes	Cyprinidae	*Poropuntius*	Cyprinid fish								1/3			3/4
Cypriniformes	Cyprinidae	*Poropuntius*	Cyprinid fish								3/3			
Cypriniformes	Cyprinidae	*Cyprinus carpio*	Common carp			1/1	4/7		1/3	2/4		1/6		
Cypriniformes	Cyprinidae	*Thynnichthys polylepis*	Bauk ketuk / Bauk pipih							3/4	3/3			4/4
Cypriniformes	Cyprinidae	*Hampala*	Cyprinid fish									1/6		

				BA	BE	BU	LT	TO	RP	SE	ME	RK	MA	CH
Cypriniformes	Cyprinidae	*Labiobarbus*	Cyprinid fish											2/4
Cypriniformes	Cyprinidae	*Mystacoleucus*	Cyprinid fish											3/4
Cypriniformes	Cyprinidae	*Neolissochilus soroides*	Soro Brook Carp											2/4
Cypriniformes	Cyprinidae	*Osteochilus*	Cyprinid fish								1/3			
Cypriniformes	Cyprinidae	*Osteochilus*	Cyprinid fish				1/7							
Cypriniformes	Cyprinidae	*Osteochilus*	Cyprinid fish		3/3		2/7		2/3	2/4	3/3			
Cypriniformes	Cyprinidae	*Osteochilus*	Cyprinid fish											3/4
Cypriniformes	Cyprinidae	*Osteochilus waandersii*	Kepiat / Pahat / Umpan							4/4	1/3			4/4
Cypriniformes	Cyprinidae	*Rasbora*	Cyprinid fish		1/3		1/7							
Cypriniformes	Cyprinidae	*Rasbora*	Cyprinid fish							2/4				
Cypriniformes	Cyprinidae	*Rasbora*	Cyprinid fish											1/4
Cypriniformes	-	-	-				7/7							
Cypriniformes	-	-	-											2/4
Cyprinodontiformes	Aplocheilidae	*Aplocheilus panchax*	Blue Panchax				1/7							
Cyprinodontiformes	Poeciliidae	*Gambusia affinis*	Western Mosquitofish				1/7							
Cyprinodontiformes	Poeciliidae	*Poecilia*	Guppy sp.		2/3									
Cyprinodontiformes	Poeciliidae	*Xiphophorus hellerii*	Green Swordtail		2/3	1/1	4/7							
Perciformes	Ambassidae	-	Asiatic glassfish					10/10						
Perciformes	Gobiidae	*Gobiopterus*	Goby						1/3					
Perciformes	Gobiidae	*Glossogobius*	Goby										1/3	
Perciformes	Gobiidae	*Glossogobius*	Goby										1/3	
Perciformes	Oxudercidae	*Mugilogobius*	Goby										1/3	
Perciformes	Eleotridae	*Oxyeleotris marmorata*	Marble Goby					1/10						
Perciformes	Gobiidae	*Pseudogobiopsis oligactis*	Bigmouth Stream Goby											1/4
Gonorynchiformes	Chanidae	*Chanos chanos*	Milkfish				2/7							
Osteoglossiformes	Notopteridae	*Chitala lopis*	Giant Featherback											2/4

				BA	BE	BU	LT	TO	RP	SE	ME	RK	MA	CH
Osteoglossiformes	Notopteridae	*Notopterus notopterus*	Bronze Featherback						1/3					
Scombriformes	Scombridae	*Auxis*	Tuna										1/4	
Siluriformes	Siluridae	-								1/3				
Siluriformes	Bagridae	*Hemibagrus*									1/3			3/4
Siluriformes	Bagridae	*Mystus*	Catfish sp							1/4	3/3			
Siluriformes	Bagridae	*Mystus*	Catfish sp											2/4
Siluriformes	Clariidae	*Clarias batrachus*	Walking Catfish		2/3		1/7						1/4	
Siluriformes	Loricariidae	*Pterygoplichthys pardalis*	Amazon Sailfin Catfish				1/7							
Siluriformes	Pangasiidae		Shark catfish sp							1/4				
Siluriformes	Pangasiidae	-	Shark catfish sp											2/4
Synbranchiformes	Mastacembelidae	-	Spiny eel fish sp								1/3			
Primates	Cercopithecidae	*Presbytis melalophos*	Mitred Leaf Monkey											1/4
			Total OTUs	3	10	5	27	6	11	19	18	5	11	27

5.4.2 16S Marker

There were 50 high quality OTUs produced after bioinformatic filtering, of which 19 could be assigned to species level, and 36 could be assigned to genus level, from eight orders. Taxonomic orders consisted of the fish groups Cypriniformes (33) and Cyprinodontiformes (2); the amphibian group Anura (2); and the mammal groups Carnivora (4), Chiroptera (1), Primates (2), Rodentia (1), Ruminantia (3), and Suina (1). OTUs which could be assigned to species included a range of fish, amphibians and mammals.

The Chenderoh Reservoir in Malaysia and Lake Laut Tawar in Sumatra had the highest OTU richness (20), whilst Lake Batur in Bali and Lake Matano in Sulawesi had the lowest (0).

As discussed in Chapter 4, OTU 7, assigned to *Bos taurus* (Cattle) was removed from the metabarcoding data from this analysis. This OTU matched with 100% Query Cover and Identity to the Cattle (*Bos Taurus*), the Zebu (*Bos indicus*), the Aurox (*Bos primigenius*), and the worm (*Phascolosoma esculenta*) (a species commonly used for biological derivatives in biochemical research e.g. Wu *et al*. (2014). Although this is likely derived from true cattle eDNA, it may either be a Genbank error or possibly an amplification of Bovine serum albumin (BSA) used in the PCR set up.

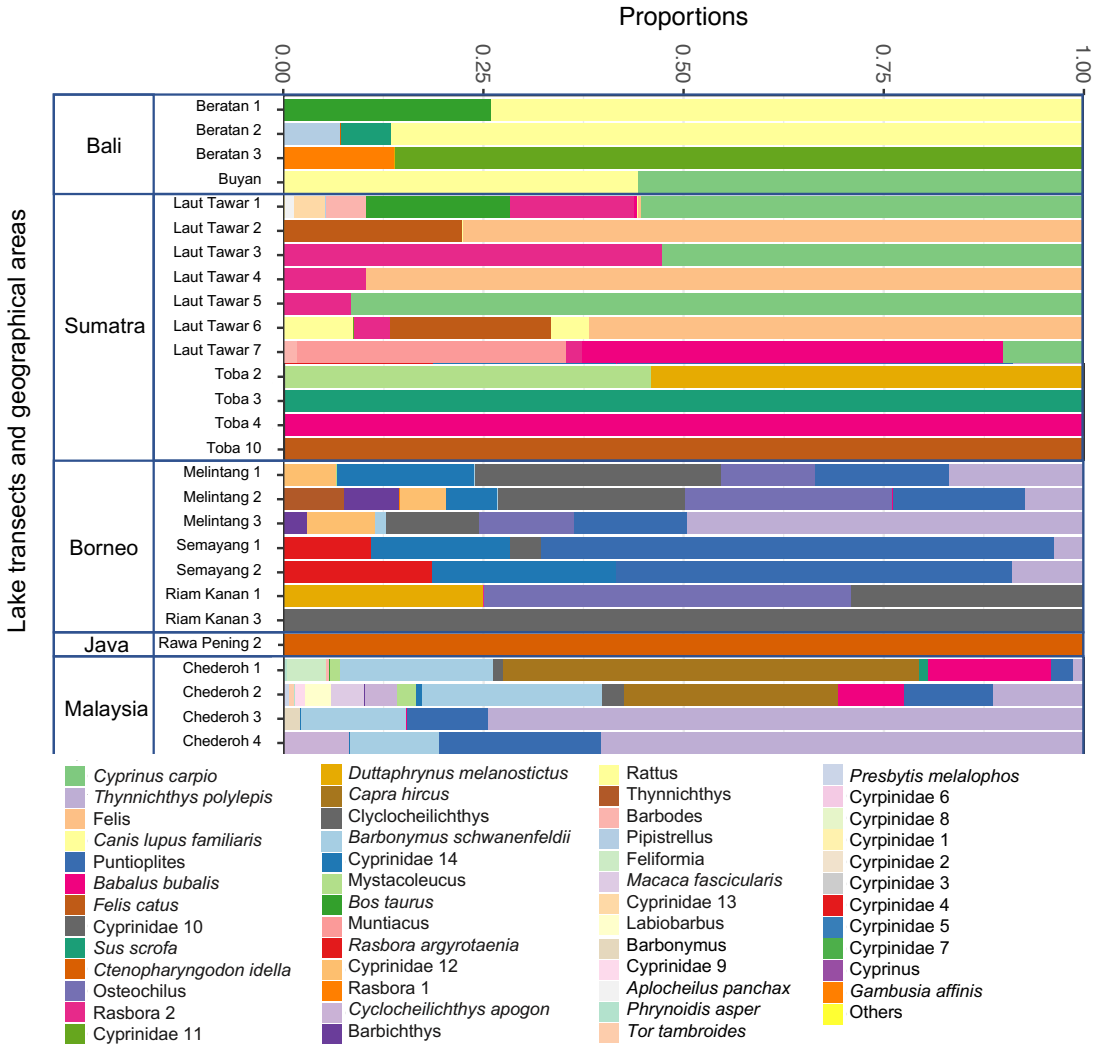


Figure 5.17. Taxa plot of 16S marker reads across all lakes. This plot was created using the top 50 most abundant taxa selected from the species level and higher depending on the filtering criteria. Several of the transects contained no 16S reads after filtering and so were removed from the taxa plot. These were: Batur 1, 2, Matano 1, 2, 3, 4, Riam Kanan 2, 4, 5, 6, Rawa Pening 1, 3, Semayang 3, 4, Toba 1, 5, 6, 7, 8, 9.

Table 5.5. Taxonomic assignments of 16S OTUs, and their presence per lake. The number of transects with a positive detection are shown in each coloured cell. Colour coding is based on the total reads found from all samples from one lake after bioinformatic filtering but before normalisation. Red = > 10,000 reads across a single, orange = 5,000 – 9,999, yellow = 2,500 – 4,999, green = 1,000 – 2,499 and blue = 20 – 999. BA = Lake Batur, BE = Lake Beratan, BU = Lake Buyan, TO = Lake Toba, RP = Lake Rawa Pening, SE = Lake Semayang, ME = Lake Melintang, RK = the Riam Kanan Reservoir, MA = Lake Matano, CH = the Chenderoh Reservoir. Species with a * were accepted beneath the 90 - 98% identity threshold as they are the only species within the genus occurring in this geographic region

16S														
Order	**Family**	**Genus / Species**	**Common Name**	**BA**	**BE**	**BU**	**LT**	**TO**	**RP**	**SE**	**ME**	**RK**	**MA**	**CH**
Cypriniformes	Cyprinidae	*Barbichthys laevis**	Sucker Barb								2/4			
Cypriniformes	Cyprinidae	*Barbodes*	Cyprinid fish				2/7							1/4
Cypriniformes	Cyprinidae	*Barbonymus schwanenfeldii*	Tinfoil Barb								1/3			4/4
Cypriniformes	Cyprinidae	*Barbonymus*	Cyprinid fish											1/4
Cypriniformes	Cyprinidae	*Ctenopharyngodon idella*	Grass Carp						1/3					
Cypriniformes	Cyprinidae	*Cyclocheilichthys*	Cyprinid fish							1/3	3/3			2/4
Cypriniformes	Cyprinidae	*Cyclocheilichthys apogon*	Beardless Barb											2/4
Cypriniformes	Cyprinidae	*Cyprinus carpio*	Common Carp			1/1	4/7							
Cypriniformes	Cyprinidae	*Tor tambroides*	Mahseer											1/4
Cypriniformes	Cyprinidae	*Labiobarbus*	Cyprinid fish											1/4
Cypriniformes	Cyprinidae	*Mystacoleucus*	Cyprinid fish					1/10						2/4
Cypriniformes	Cyprinidae	*Osteochilus*	Cyprinid fish				1/7				3/3	1/6		
Cypriniformes	Cyprinidae	*Puntioplites*	Cyprinid fish							2/3	3/3			4/4
Cypriniformes	Cyprinidae	*Rasbora argyrotaenia*	Silver Rasbora							2/3				
Cypriniformes	Cyprinidae	*Rasbora*	Cyprinid fish		1/3		1/7							
Cypriniformes	Cyprinidae	*Rasbora*	Cyprinid fish				6/7							
Cypriniformes	Cyprinidae	*Thynnichthys*	Cyprinid fish								1/3			
Cypriniformes	Cyprinidae	*Thynnichthys polylepis*	Bauk ketuk / Bauk pipih								3/3			
Cypriniformes	Cyprinidae	-	Cyprinid fish				1/7							
Cypriniformes	Cyprinidae	-	Cyprinid fish				1/7							
Cypriniformes	Cyprinidae	-	Cyprinid fish				1/7							
Cypriniformes	Cyprinidae	-	Cyprinid fish				1/7							
Cypriniformes	Cyprinidae	-	Cyprinid fish				1/7							
Cypriniformes	Cyprinidae	-	Cyprinid fish				1/7							
Cypriniformes	Cyprinidae	-	Cyprinid fish				1/7							

Order	**Family**	***Genus / Species***	**Common Name**	**BA**	**BE**	**BU**	**LT**	**TO**	**RP**	**SE**	**ME**	**RK**	**MA**	**CH**
Cypriniformes	Cyprinidae	-	Cyprinid fish											1/7
Cypriniformes	Cyprinidae	-	Cyprinid fish											1/4
Cypriniformes	Cyprinidae	-	Cyprinid fish									2/6		
Cypriniformes	Cyprinidae	-	Cyprinid fish		1/3									
Cypriniformes	Cyprinidae	-	Cyprinid fish								3/3			
Cypriniformes	Cyprinidae	-	Cyprinid fish				1/7							
Cypriniformes	Cyprinidae	-	Cyprinid fish							2/4	2/3			1/4
Cyprinodontiformes	Aplocheilidae	*Aplocheilus panchax*	Blue Panchax				1/7							
Cyprinodontiformes	Poeciliidae	*Gambusia affinis*	Mosquitofish				1/7							
Anura	Bufonidae	*Duttaphrynus melanostictus*	Asian Common Toad					1/10				1/7		
Anura	Bufonidae	*Phrynoidis asper*	Asian Giant Toad											2/4
Carnivora	Canidae	*Canis lupus familiaris*	Domestic Dog		2/3	1/1	2/7							
Carnivora	Felidae	*Felis catus*	Domestic Cat				2/7	1/7						
Carnivora	Felidae	*Felis catus*	Domestic Cat				4/7							
Carnivora	Feliformia	-	Cat-like carnivore											1/4
Chiroptera	Vespertilionidae	*Pipistrellus*	Bat		1/3									
Primates	Cercopithecidae	*Macaca fascicularis*	Crab-eating macaque											1/4
Primates	Cercopithecidae	*Presbytis melalophos*	Mitred Leaf Monkey											1/4
Rodentia	Muridae	*Rattus*	Rat											1/4
Ruminantia	Bovidae	*Bubalus bubalis*	Water Buffalo				2/7	1/10						2/4
Ruminantia	Bovidae	*Capra hircus*	Goat											2/4
Ruminantia	Cervidae	*Muntiacus*	Mutjac				1/7							
Suina	Suidae	*Sus scrofa*	Wildboar		1/3			1/10						1/4
			Total OTUs	0	5	2	20	5	1	4	9	3	0	20

5.4.3 COI marker

The COI metabarcoding data created using the 313 bp fragment by Leray *et al*. (2013) was dominated by microfauna, meiofauna and microalgae, and so based on these data is not recommended for the monitoring of vertebrates through aquatic eDNA. There were eight OTUs accepted to species level: *Diaphanosoma excisum* (a species of freshwater ctenopod water flea in the family Sididae) *Helostoma temminkii* (Kissing Gourami: a common Southeast Asian aquaculture fish species,), *Selar crumenophthalmus*, (Bigeye Scad: a marine fish species), *Brachionus calyciflorus* (a freshwater planktonic rotifer species), *Euchlanis dilatata* (another freshwater planktonic rotifer species), *Eodiaptomus wolterecki* (a freshwater copepod zooplankton, containing two different OTUs) and *Sinanodonta woodiana* (Chinese Pond Mussel).

The Chinese Pond Mussel (*Sinanodonta woodiana*) is native to East Asia, but is an introduced species in the Indonesian islands of the Malay Archipelago (Bolotov *et al*. 2016) likely through the ornamental pet trade (Ng *et al*. 2015). This species was found from the COI metabarcoding data presented here from Lake Laut Tawar in Aceh, Sumatra, where according to Bolotov *et al*. (2016), it has not been previously recorded, although it was not detected from Lake Beratan where it has been previously recorded (Whitten *et al*. 1996).

The *Rasbora* OTU found from the COI metabarcoding data from Lake Laut Tawar matched with 100% query cover and 100% identity to a sequence which that the BLAST matched to "Rasbora sp. ZAM-2010 voucher R3" and "Rasbora sp. ZAM-2010 voucher R2" from a study from 2013 investigating the different *Rasbora* fish of Lake Laut Tawar, which had all been classified as *Rasbora tawarensis* (Muchlisin, 2013). The Bahasa Indonesia names for these three fish are Depik, Eos and Relo, which local fisherman categorised based on size. Genetic investigation suggested that Depik and Eos were in fact variations of *Rasbora tawarensis*, whilst Relo is another separate cryptic species (Muchlisin, 2013), which the 313 bp barcode from the data herein matched to perfectly. This highlights the need for more molecular barcoding of Indonesian fish to populate genetic records for biodiversity and fisheries monitoring.

The zooplankton species *Eodiaptomus wolterecki* is native to the ancient lakes of Eastern Sulawesi (Sabo *et al*. 2008), and was only detected in the data herein from samples from Lake Matano, Sulawesi. This supports the reliability of the metabarcoding approach used within this study.

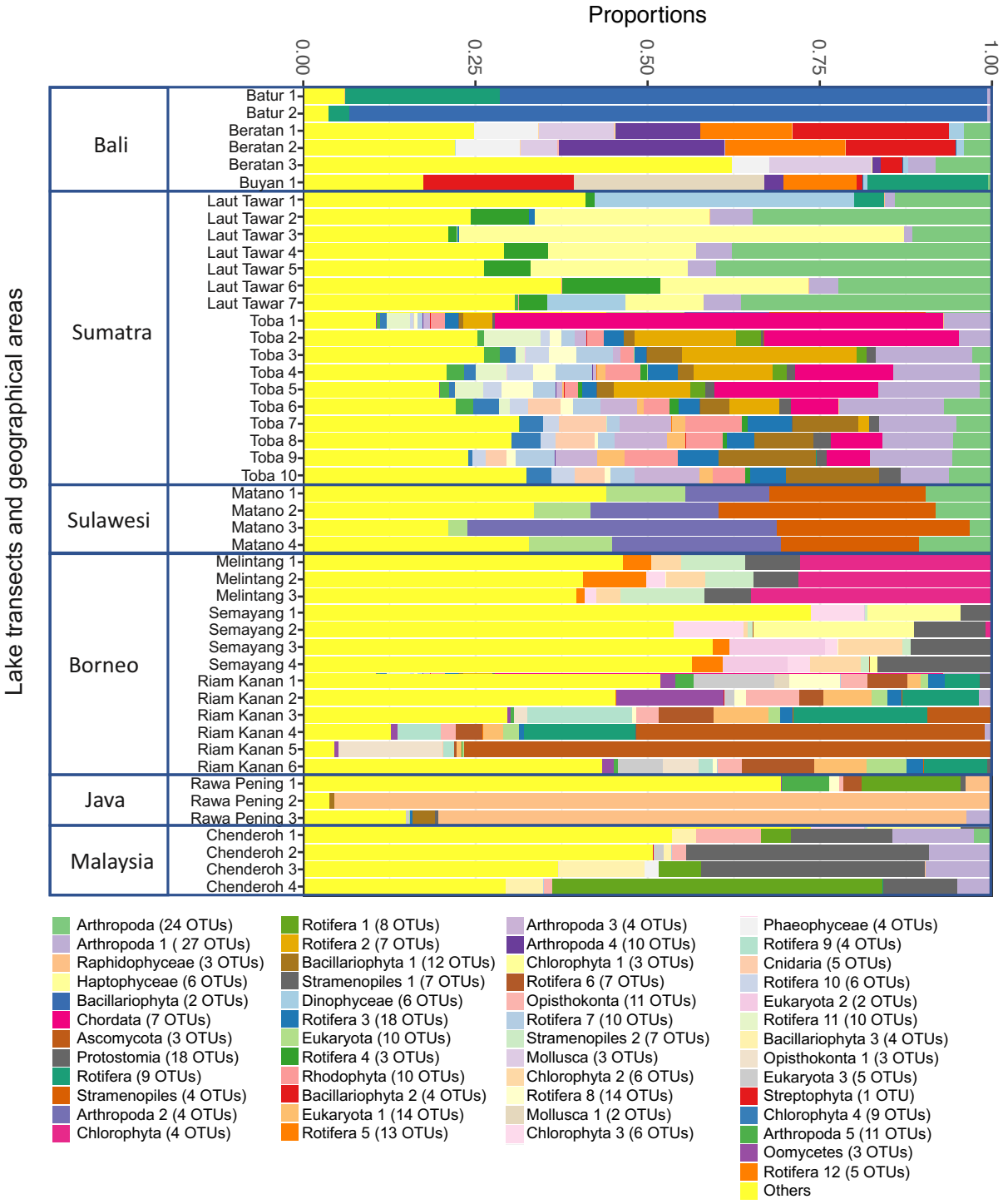


Figure 5.18. Taxa plot of COI marker reads across all lakes. This plot was created using the top 50 most abundant taxa selected from the phylum level and higher. The reads placed into 'Other' consisted of 1073 OTUs within the phyla Annelida, Arthropoda, Ascomycota, Bacillariophyta, Bicosoecida, Bilateria, Blastocladiomycota, Chordata, Chlorophyta, Chrysophyceae, Cryptophyta, Choanozoa, Cnidaria, Collodictyonidae, Dinophyceae, Eukaryota, Eumetazoa, Fungi, Jakobida, Metazoa, Mollusca, Ochrophyta, Oomycetes, Opisthokonta, Phaeophyceae, Platyhelminthes, Porifera, Proteobacteria, Protostomia, Raphidophyceae, Rhodophyta, Rotifera, Stramenopiles and Streptophyta. Each lake shows a similar community composition, with some phyla such as Bacillariophyta and Streptophyta only found in specific lakes.

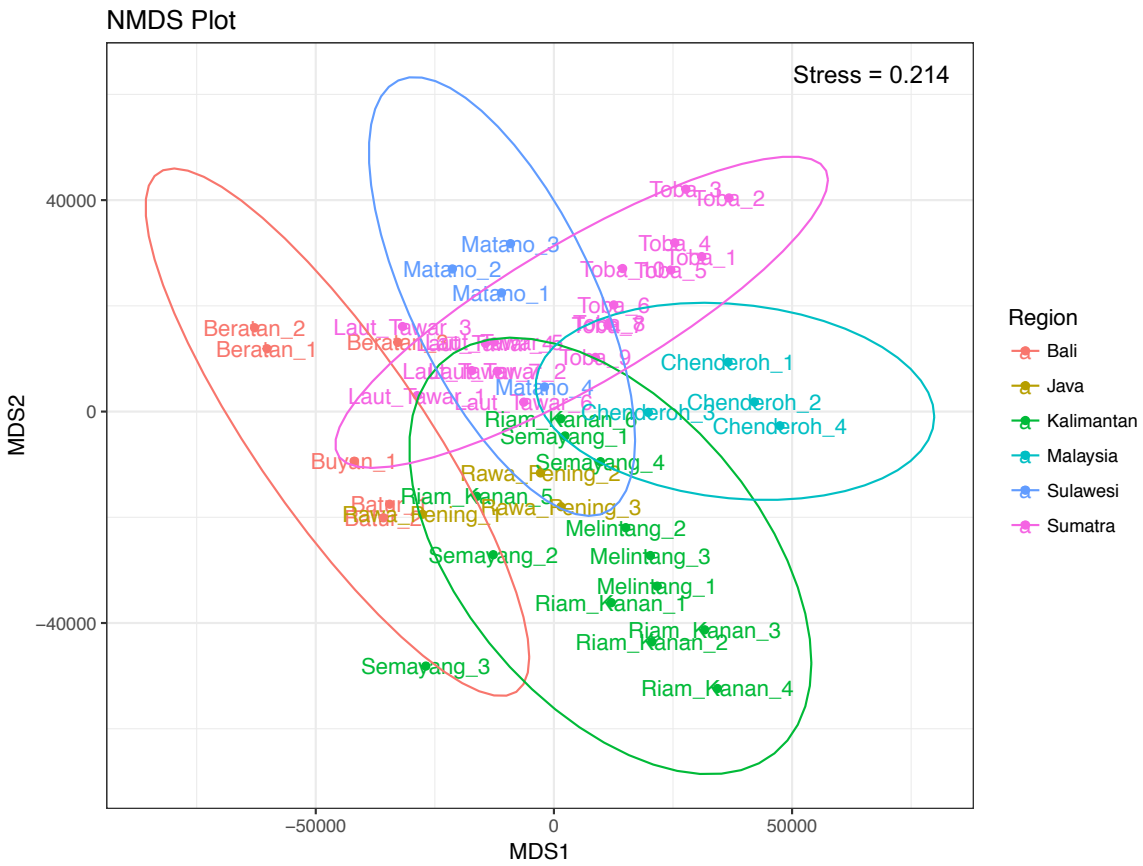


Figure 5.19. NMDS Plot of OTUs per region. A Non-Metric Multidimensional Scale Plot (NMDS) created using the normalised read counts from the combined OTU tables from the COI, 12S and 16S markers, grouped according to region (stress = 0.214).

The distance matrix derived from the Manhattan method showed a statistical impact of Region (Adonis PERMANOVA, $R^2 = 0.407$; $P = 0.001$) on OTU community composition. Generally, the lakes from individual regions clustered together, with some overlap. The only lake from Java (Rawa Pening) had too few points to create an individual cluster, but points were clustered with most overlap with the Kalimantan group. The Balinese lakes (Beratan, Buyan, Batur) clustered together (orange), and were furthest away from the Chenderoh samples (Malaysia). The Kalimantan (green) lakes also clustered together (Melintang, Semayang and Riam Kanan). The two lakes from Sumatra (Toba and Laut Tawar) showed unique spatial clustering, as did the samples from Chenderoh (Malaysia, turquoise) and from Matano (Sulawesi, blue).

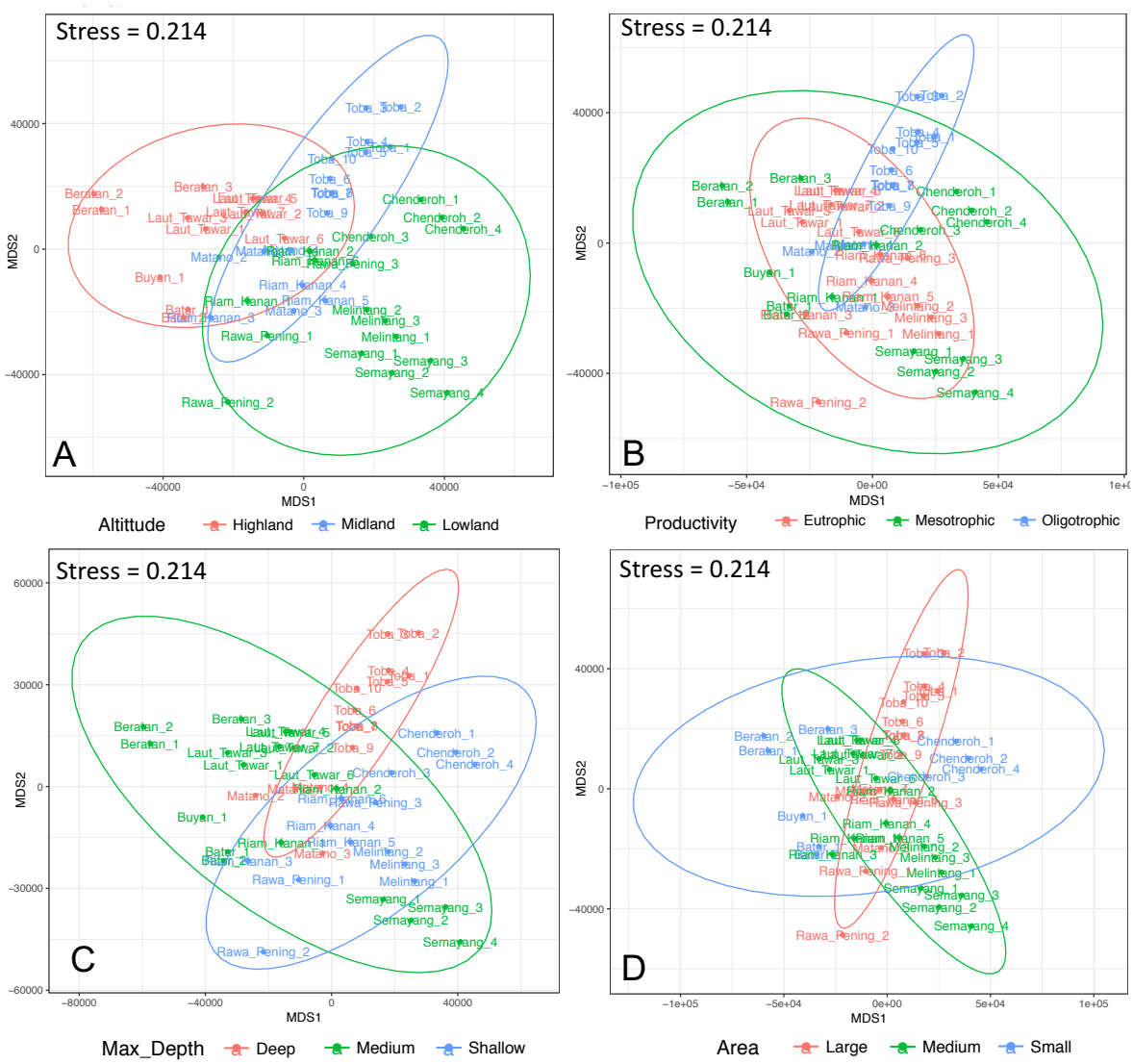


Figure 5.20. NMDS plots of OTU community composition according to habitat variables. Altitude, Productivity, Max Depth and Area. A Non-Metric Multidimensional Scale Plot (NMDS) created using the normalised read counts from the combined OTU tables from the COI, 12S and 16S markers, grouped according to region (stress = 0.214).

The most defined clusters according to a particular variable are seen in the NMDS plot of OTU community by Altitude (Figure 5.20 A), in which the highland lakes cluster at the left (red), followed by midland lakes in the centre (blue), and lowland lakes at the right (green).The distance matrix derived from the Manhattan method showed a statistical impact of all habitat variables on OTU community composition, Altitude (Adonis PERMANOVA, R^2 = 0.219; P = 0.001); Productivity (Adonis PERMANOVA, R^2 = 0.185; P =0.001); Max Depth (Adonis PERMANOVA, R^2 = 0.201; P =0.001) and Area (Adonis PERMANOVA, R^2 = 0.178; P =0.001).

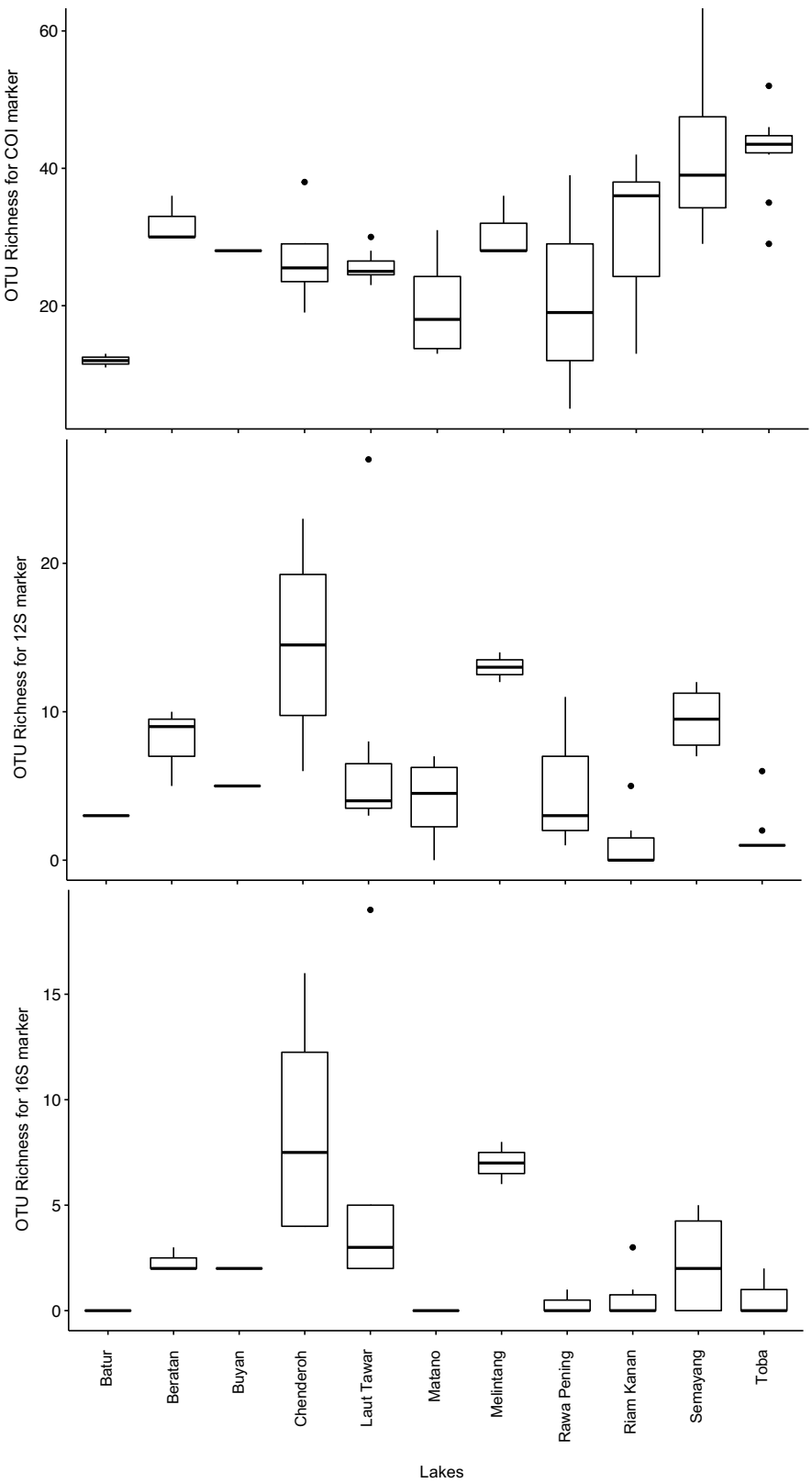


Figure 5.21 OTU richness per lake by specific markers. OTU richness by the COI marker (top), 12S marker (middle) and 16S marker (bottom).

The ANOVA comparing OTU richness between lakes showed a significant difference for each marker combination used (all markers, P = 0.00234; 12S marker, P = 0.0008, 16S marker, P = 0.001, and COI marker, P = 0.0002).

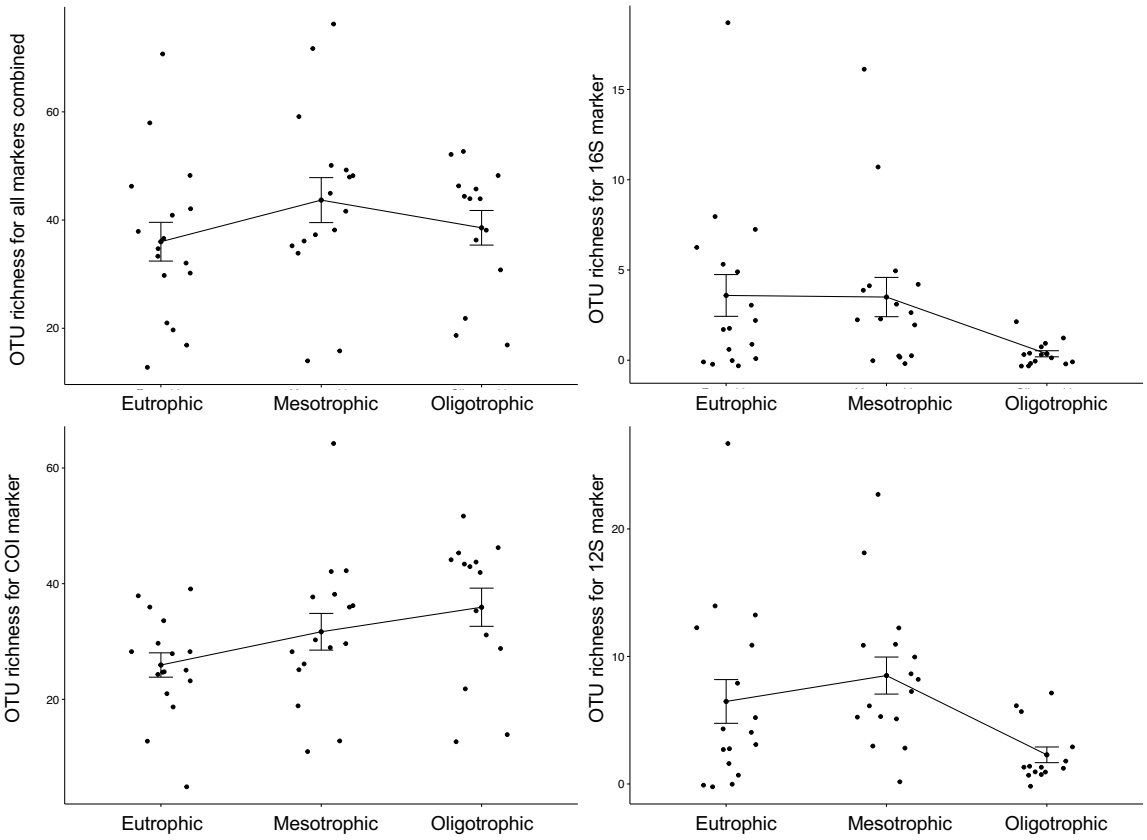


Figure 5.22. OTU Richness by productivity. Lakes were grouped into 'eutrophic', 'mesotrophic' or 'oligotrophic' based on the description from the literature, and the OTU richness compared.

When comparing OTU richness between groups based on productivity (eutrophic, mesotrophic and oligotrophic) there was no difference when analysing all markers combined (P = 0.323) or for COI alone (P = 0.056), but there was a significant difference between groups for 12S (P = 0.013) and for 16S (0.043). There was a lower OTU richness for 12S and 16S in oligotrophic lakes, and a higher OTU richness in eutrophic or mesotrophic lakes. This is the expected pattern, as oligotrophic lakes are less nutrient dense, and so support less plant life and subsequent succession of biodiversity.

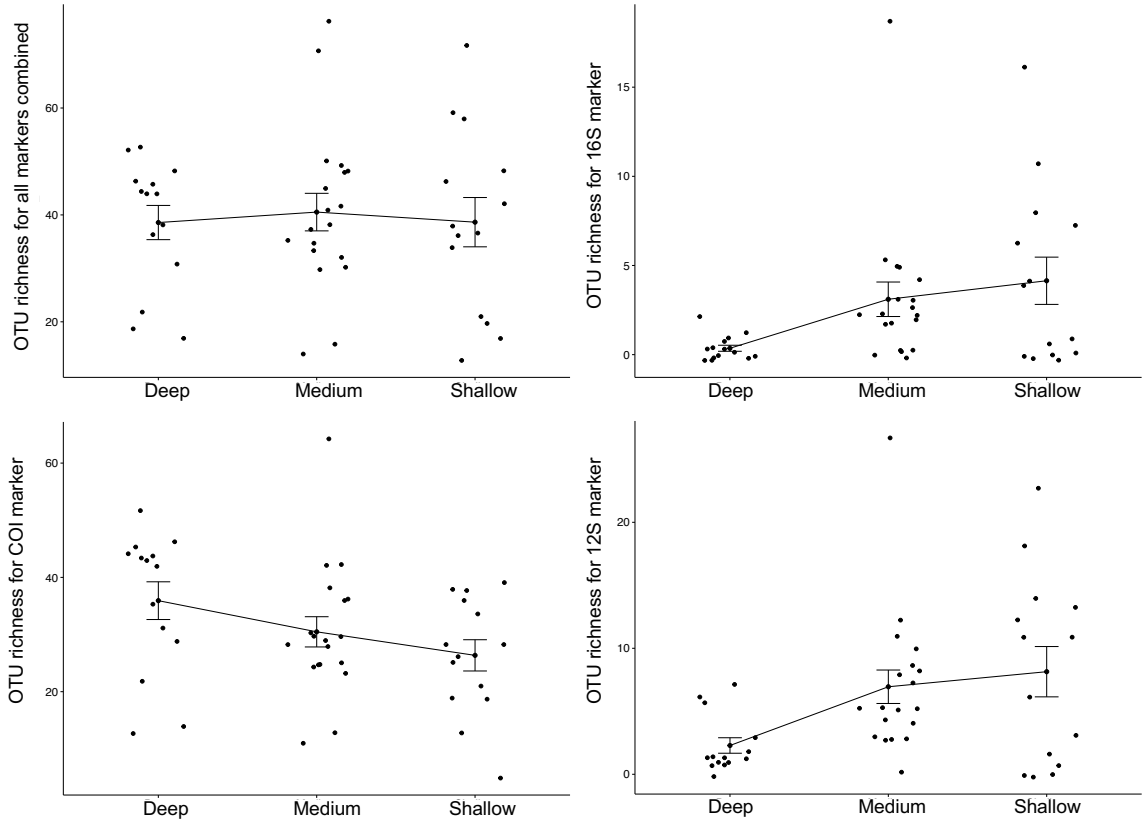


Figure 5.23. OTU Richness by maximum depth. Lakes were grouped into 'deep, 'medium or 'shallow' based on the maximum depths described from the literature, and the OTU richness compared. Deep lakes were those with a maximum depth of more than 100 m (n = 2 lakes, 14 transects), medium lakes were those with a maximum depth of 21 – 100 m (n = 5 lakes, 19 transects), and shallow lakes were those with a maximum depth between 0 – 20 m (n = 4 lakes, 14 transects).

The ANOVA comparing OTU richness according to lake depth showed no significant difference when using all markers combined (P = 0.912), but a significant difference when only using the 12S marker (P = 0.019), the COI marker (P = 0.095), and the 16S marker (P = 0.0326). For the COI marker, deep lakes had a greater OTU richness, and for the 12S and 16S marker, shallow lakes had a greater OTU richness.

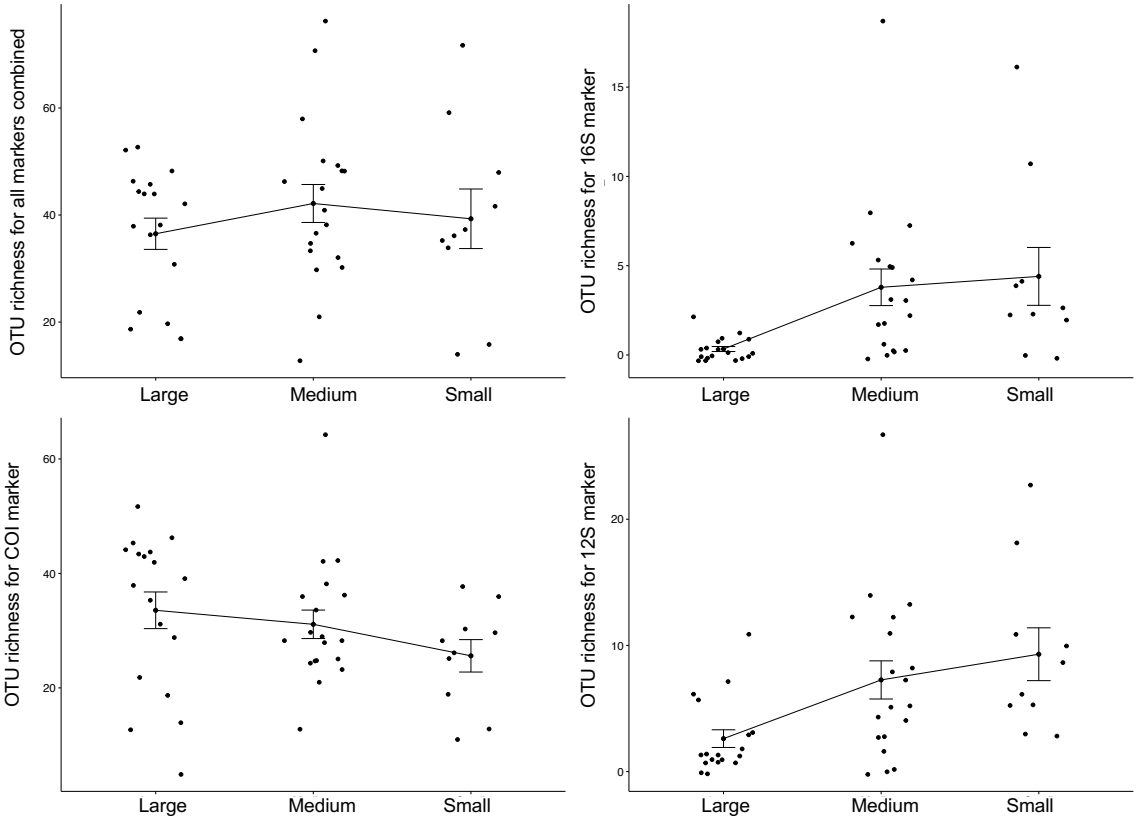


Figure 5.24 OTU Richness by area. Lakes were grouped into 'large, 'medium or 'small based on the area described from the literature, and the OTU richness was compared. Large lakes were those with an area of more than 100 km^2 (n = 3 lakes, 18 transects), medium lakes were those with an area 11 – 100 km^2 (n = 4 lakes, 19 transects), and small lakes were those with a maximum depth between 0 – 10 km^2 (n = 4 lakes, 10 transects).

In tropical Asian lakes, fish species richness is mostly predicted by lake area rather than other variables which predict fish species richness in temperate lakes (Amarasinghe and Welcomme, 2002). There was no significant difference between area categories when analysing all markers combined (P = 0.516) or COI alone (P = 0.233), but there was when analysing the 12S marker (P = 0.006) and 16S marker (P = 0.007) alone. There was a higher OTU richness in small and medium sized lakes compared to large lakes. This may be due to the fact that there are interacting factors within this dataset, such as large lakes (Lake Toba and Matano) also being oligotrophic.

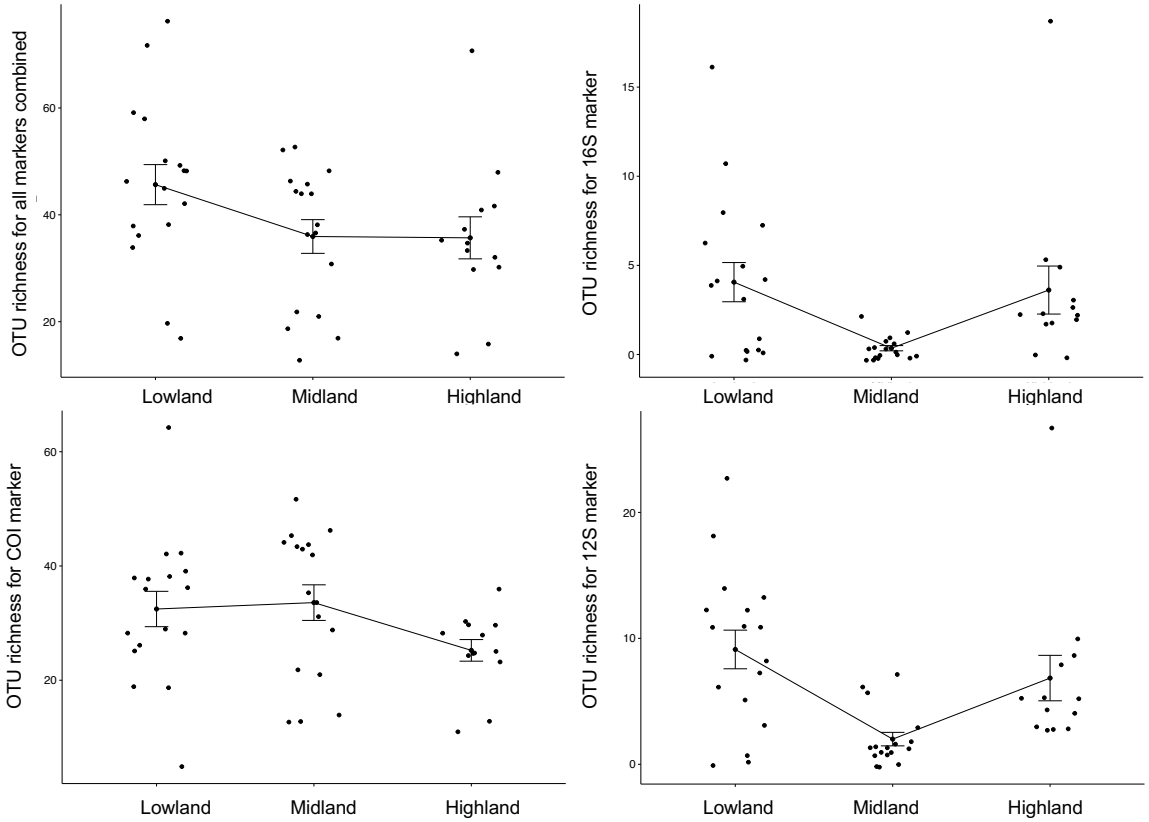


Figure 5.25 OTU Richness by altitude. Altitude' was the height above sea level at which the lake resided, either 'highland' (altitude of > 1000 m above sea level, n = 4 lakes, 13 transects), 'midland' (altitude of > 101 - 999 m above sea level, n = 3 lakes, 17 transects) or 'lowland' (altitude of 0 – 100 m above sea level, n = 4 lakes, 17 transects).

The ANOVA comparing OTU richness according to altitude showed no significant difference when using all markers combined (P = 0.088), or for the COI marker alone (P = 0.122), but a significant difference when only using the 12S marker (P = 0.001), and the 16S marker (P = 0.012). For the 12S and 16S marker, lowland lakes had a slightly greater OTU richness, although interestingly there was a higher OTU richness observed at either end of the altitude scale, with medium altitude lakes showing a lower OTU richness.

5.5 Discussion

This study illustrates the success of aquatic eDNA metabarcoding for the detection of the vertebrate species of Southeast Asia. Many fish and mammal species were detected from relatively few samples at sites which mostly have no temporal replicates. The combination of markers used yielded a variety of taxonomic information, and was an important factor in detecting the range of fish species observed.

5.5.1 Fish species

After filtering, the 12S marker detected almost entirely fish species, and also one mammal species. Of the fish species detected from all lakes, many were either important fishery species or grown for commercial aquaculture, as is expected due to the high density of stocks of these species within lakes. In addition, several invasive species were detected, as well as some rare native fish. Several species amplified using the 12S marker were also amplified using the 16S marker, creating a higher level of confidence in these assignments.

5.5.2 Native fisheries and aquaculture species

Many OTUs were assigned to species which are important to either local fisheries, commercial aquaculture, or sport fishing. Native fish detected which are used for these purposes include the Climbing Perch (*Anabas testudineus*), the Giant Snakehead (*Channa micropeltes*), the Kissing Gourami (*Helostoma temminckii*), the Three-Spot-Gourami (*Trichopodus trichopterus*), and the Common Carp (*Cyprinus carpio*).

The OTU assigned to an *Anabas* species (12S) is very likely to be the Climbing Perch (*Anabas testudineus*), as this is the only species of this genus occurring in Southeast Asia. There are only two species within this genus, and only one of them *A. testudineus* occurs in Malaysia and Indonesia, whilst the other, *A. cobojius* is native to India and Bangladesh.

The Giant Snakehead (*Channa micropeltes*), is known from West Kalimantan, but no literature was found describing this species from East Kalimantan where it was detected from the eDNA samples (12S). It could however, have already existed unrecorded in the Semayang / Melintang region, or been introduced from commercial or gamefish fisheries. This species is however known from the Chenderoh Reservoir, but was not detected there from this data (Kah-Wai and Ali, 2000; Hashim *et al*. 2012).

The Kissing Gourami (*Helostoma temminckii*) is known from Semayang and Melintang where it was detected from the 12S data, and is native to Indonesia but commonly

used in aquaculture (Haryono, 2006). The OTU (from the 12S data) assigned to the genus *Trichopodus* first matched to *Trichogaster*, the old name for *Trichopodus*. This is likely to be the Three-Spot-Gourami, *Trichopodus trichopterus* (previously *Trichogaster trichopterus)*, a common fisheries species found in Melintang (Haryono, 2006), Semayang (Haryono, 2006), Laut Tawar (Muchlisin *et al*. 2009), Toba (Fishbase, 2017a), and Chenderoh (Hashim *et al*. 2012). The Common Carp (*Cyprinus carpio*) was detected in five out of eleven lakes (five from 12S and two from16S data), four of which are lakes where this species had been previously recorded - Buyan (Restu *et al*. 2016; Green *et al*. 1978), Laut Tawar (Muchlisin *et al*. 2009; Muchlisin and Azizah, 2009; Muchlisin *et al*. 2010; Muchlisin, 2012), Riam Kanan (De Silva, 1987) and Rawa Pening (Hutarabat *et al*. 1986). The Common Carp was also detected from Lake Semayang, although no mention of this could be found in the literature. In addition, this species had been previously recorded from other lakes (Batur, Beratan, Matano and Toba), but was not detected from this eDNA survey (Green *et al*. 1978; Sentosa *et al*. 2013; Whitten *et al*. 1996; Versteegh, 2010; Wijopriono *et al*. 2010.

The 16S data detected the species *Tor tambroides* through the BLAST assignment by MEGAN, however this is actually a misidentification according to Fishbase, and in fact refers to *Tor tambra*.

Two OTUs from the *Cyclocheilichthys* genus were detected. In the Chenderoh Reservoir, these were the Beardless Barb *Cyclocheilichthys apogon* (Kah-Wai and Ali, 2000; Hashim *et al*. 2012) and another unknown species, possibly *C. armatus* (Hashim *et al*. 2012) or *C. heteronema* (Kah-Wai and Ali, 2000; Hashim *et al*. 2012)

5.5.3 Endemic or rare native fish species

The eDNA samples (12S) detected the Sailfin Silversides endemic to the Malili Lake system (Herder *et al*. 2008). These included the 'Roundfin' Sailfin Silverside fish, *Telmatherina prognatha* which is endemic to Lake Matano (Kurniawan and Subehi, 2016), the only lake from which the OTU assigned to this species was amplified. Another OTU from the Telmatherinidae family was amplified from the Lake Matano samples, which could be one of nine potential endemic species (Fishbase, 2017b). The Perak River Sprat (*Clupeichthys perakensis*) is native to the Perak River which flows through the Chenderoh Reservoir (Whitehead, 1985), the lake from which this OTU was detected from the 12S data. Two OTUs assigned to the *Poropuntius* genus (from the 12S data) were detected from Melintang and Chenderoh. One of these at least is likely to be Waander's Bony Lipped Barb

Poropuntius deauratus (Hashim *et al*. 2012), although there are three other *Poropuntius* species in Indonesia, and two others in Malaysia (Fishbase, 2017b).

Two OTUs were detected (12S) from the *Pristolepis* genus, one from the Semayang – Melintang system, and another from the Chenderoh Reservoir. This is likely to be *Pristolepis fasciata*, the only species within this genus recorded from both Chenderoh (Kah-Wai and Ali, 2000; Hashim *et al*. 2012) and the Semayang – Melintang system (Haryono, 2006). However, as two different OTUs were observed, divided between East Kalimantan and Malaysia, these may be different species, or the same species with distinct haplotypes.

Two previously recorded 'Barb' species of Cyprinid fish were detected from the Chenderoh Reservoir alone - the Beardless Barb (*Cyclocheilichthys apogon*) (16S) (Kah-Wai and Ali, 2000; Hashim *et al*. 2012), and the Silver Barb (*Barbonymus gonionotus*) (12S) (Kah-Wai and Ali, 2000; Hashim *et al*. 2012). The Sucker Barb (*Barbichthys laevis*) (12S and 16S) was detected from the Chenderoh Reservoir (Hashim *et al*. 2012), Semayang (Haryono, 2006; Kurniawan and Subehi, 2016) and Melintang (Haryono, 2006). The Tinfoil Barb (*Barbonymus schwanenfeldii*) was detected from the Chenderoh Reservoir (Kah-Wai and Ali, 2000; Hashim *et al*. 2012), Melintang and Semayang (Kurniawan and Subehi, 2016), through the 12S and 16S data, although this species was missing from Toba where it has been previously recorded (Fishbase 2017a).

A species of the *Crossocheilus* genus known as 'algae eaters' was also found in the Chenderoh Reservoir (12S), which could be one of seven species recorded from Malaysia. *Thynnichthys polylepis* was detected from the 12S and 16S data from a combination of Semayang, Melintang and Chenderoh with 100% Query Cover and Identity, although it is *Thynichthys vaillanti* that has been previously recorded from the Semayang – Melintang lakes (Haryono, 2006) and *Thynnichthys thynnoides* from Chenderoh (Kah-Wai and Ali, 2000; Hashim *et al*. 2012). This could be a misidentification of the species uploaded to BLAST, a misidentification of the fish recorded in the visual survey, or it may be that both species occur within this habitat.

The *Hampala* species detected from Riam Kanan (12S) is likely to be the previously recorded Hampala Barb (*Hampala macrolepidota*) (Hardjamulia and Suwignyo, 1988). The Labiobarbus species from the Chenderoh Reservoir (12S and 16S) could be either *Labiobarbus fasciatus* (Hashim *et al*. 2012), *Labiobarbus leptocheilus* (Kah-Wai and Ali, 2000), or *Labiobarbus lineatus* (Kah-Wai and Ali, 2000; Hashim *et al*. 2012).

The *Mystacoleucus* species detected from the Chenderoh Reservoir (12S and 16S) is likely to be *Mystacoleucus marginatus*, as has previously been recorded here (Hashim *et al*.

2012). The Soro Brook Carp (*Neolissochilus soroides*), detected from the Chenderoh Reservoir (12S), was not recorded by Hashim *et al*. (2012) in their study, although it was recorded upstream in Lake Temengor. This species may have been unobserved in this 2012 study, or it may have expanded its range down the Perak River to the Chenderoh Reservoir. It is also possible that eDNA from upstream Temengor travelled down to the Chenderoh Reservoir, resulting in a positive detection without the local presence of this species.

There were four different *Osteochilus* OTUs with different distributions across lakes (12S data). The first, only found in Lake Melintang could be the only species from this genus recorded here - *Osteochilus kappenii* (Haryono, 2006), or any of the other species recorded from nearby Semayang (*Osteochilus vittatus*, *Osteochilus kelabau*, *Osteochilus melanopleurus*, or *Osteochilus repang*). The second, only detected from Laut Tawar, is likely to be *Osteochilus kahajanensis*, endemic to Laut Tawar, and the only species within this genus recorded here (Muchlisin *et al*. 2010). The third, found in Beratan, Laut Tawar, Rawa Pening, Semayang and Melintang could be the widespread *Osteochilus vitattus* (Kah-Wai and Ali, 2000; Hashim *et al*. 2012; Sentosa *et al.* 2013; Whitten *et al.* 1996; Dahruddin *et al*. 2016; Kurniawan and Subehi, 2016), which is likely to also be the Osteochilus species observed from the 16S data from Laut Tawar, Melintang and Riam Kanan. The fourth, only found from the Chenderoh Reservoir, could be *Osteochilus melanopleurus*, *Osteochilus microcephalus*, or *Osteochilus vittatus* (Kah-Wai and Ali, 2000; Hashim *et al*. 2012). The diversity and prevalence of this genus and the lack of species level assignment indicates the need for more barcoding work of cyprinid fish from Southeast Asia. The OTU assigned to *Osteochilus waandersii* (which initially matched to *Puntioplites waandersii*, the previously accepted name) was detected from Semayang, Melintang and Chenderoh where it has been previously recorded, or recorded nearby (Kurniawan and Subehi, 2016; Ikhwanuddin *et al*. 2017).

There were three OTUs assigned to the *Rasbora* genus detected from the 12S data, and two from the 16S data, also with different distributions across lakes. The first, from both the 12S and 16S data, was found only in Beratan and Laut Tawar. This could be *Rasbora baliensis* (Whitten *et al*. 1996) or the Silver Rasbora (*R. argyrotaenia*) (Sentosa *et al*. 2013). The second *Rasbora* OTU from the 12S data was only found in Semayang. This may be a yet unnamed species (Haryono, 2006), or it could be the Silver Rasbora (*R. argyrotaenia*) as assigned from the 16S data to an OTU also only found in Semayang. The third *Rasbora* OTU only found in Chenderoh (12S) may be *R. sumatrana* (Kah-Wai and Ali, 2000) or *R. tornieri* (Hashim *et al*. 2012).

The OTU assigned to the genus *Trichopsis* (12S) is likely to be *Trichopsis vittata* according to records, usually occurring in disturbed habitats such as paddy fields and ditches. It has previously been recorded from Rawa Pening where it was also detected from this eDNA data (Dahruddin *et al*. 2016).

The Blue Panchax (*Aplocheilus panchax*) was only detected from one inlet transect from Laut Tawar (12S and 16S). *A. panchax* was not recorded by a recent survey (Muchlisin, 2012), although it is a commonly observed fish across Indonesia and Malaysia. The observation of *A. panchax* and *G. affinis* in one transect of Laut Tawar supports the idea that eDNA is highly localised, as *A. panchax* was often visually observed in the shallower waters or small streams, and was not detected from the other six transects.

Three different Goby OTUs were detected from Lake Matano (12S), two *Glossogobius* and one *Mugilogobius*. These could potentially be from *Glossogobius matanensis*, *Mugilogobius latifrons* and *Mugilogobius adeia* (endemic to Lake Matano) (Nasution, 2016). Other Goby species detected included the Marble Goby (*Oxyeleotris marmorata*) from Lake Toba (12S) where it has previously been recorded (Wijopriono *et al*, 2010), and the Bigmouth Stream Goby (*Pseudogobiopsis oligactis*) from the Chenderoh Reservoir (12S) where it has been known since 1940 (Herre, 1940). The *Gobiopterus* OTU detected only from Rawa Pening (12S) is likely to be *Gobiopterus brachypterus*, the only species within this genus recorded from this lake (Dahruddin *et al*. 2016).

Two species of Notopteridae fish were detected from the 12S data, The Giant Featherback and the Bronze Featherback. These fish are important food sources (Santhanam, 2015). The native Giant Featherback (*Chitala lopis*) was only detected from the Chenderoh Reservoir where it has previously been recorded (Kah-Wai and Ali, 2000), a species also commonly caught for recreational angling. The Bronze Featherback (*Notopterus notopterus*) was only detected from Rawa Pening where it has previously been recorded (Dahruddin *et al*. 2016).

Several catfish OTUs were detected from the Semayang – Melintang lakes and the Chenderoh Reservoir. An OTU from the Siluridae family was detected from Lake Semayang only (12S), and an OTU assigned to *Hemibagrus* from Melintang and Chenderoh (12S). These are most likely to be the previously recorded native aquaculture species, *Hemibagrus nemurus* (Haryono, 2006; Hashim *et al*. 2012) and *Mystus castaneus* (Hashim *et al*. 2012). The native Walking Catfish (*Clarias batrachus*) was detected from Beratan (Whitten *et al*. 1996; Green *et al*. 1978), Laut Tawar (possibly from Muchlisin *et al*. 2010) and Matano (Herder *et al*. 2012).

Two Shark Catfish OTUs from the Pangasiidae family were detected from Semayang and Chenderoh (12S), which may be the previously recorded unknown '*Pangasius*' species from Semayang by Haryono (2006), and *Pangasius macronema* from Chenderoh by Suyatna *et al*. (2017). The BLAST result gave 100% Query Cover and Identity to *Pangasianodon hypophthalmus*, *Pangasius sutchi* and an unknown *Pangasius* species.

A Spiny Eel fish species within the Mastacembelidae family was detected from Lake Melintang only (12S). This could be (*Macrognathus aculeatus*) (Haryono, 2006) but is likely to be another species from this family instead. As the BLAST identity was only 83%, and there are three whole mitochondrial genome entries in NCBI for *M. aculeatus*, it is possible that this OTU comes from either an unknown species from the Mastacembelidae family, or from a species which does not yet have a gene reference present in NCBI.

In the case of Lake Matano, it is interesting to note that no cyprinid fish were detected. East of Wallace's Line (where Lake Matano lies), primary freshwater fishes such as cyprinids do not naturally exist, and so this was to be expected (Coates, 1985; Coates, 2002).

5.5.4 Introduced and invasive species

Other fish OTUs were detected from the eDNA samples from species which have been introduced for fisheries, aquaculture, sport fishing, ornamental or pest-control purposes. Some species have been introduced from other areas of Southeast Asia, Latin America or Africa. These include the Snakeskin Gourami (*Trichopodus pectoralis*), an *Osphronemus* species likely to be the Giant Gourami (*Osphronemus gouramy*), the Midas Cichlid (*Amphilophus citrinellus*), the Peacock Bass (*Cichla ocellaris*), the Nile Tilapia (*Oreochromis niloticus*), the Mozambique Tilapia (*Oreochromis mossambicus*), the Western Mosquitofish (*Gambusia affinis*), the Guppy (*Poecilia reticulata*), and the Green Swordtail (*Xiphophorus hellerii*).

The Snakeskin Gourami (*Trichopodus pectoralis*) was introduced from mainland Southeast Asia (the Mekong basin in Laos, Thailand, Cambodia and Vietnam) for fisheries purposes, although this has caused adverse ecological impact after introduction (Welcomme, 1988). This species was detected (12S) from Rawa Pening (Dahruddin *et al*. 2016), Melintang (Haryono, 2006), and Semayang, but not from Matano (Versteegh, 2010) and Toba (Thomas, 2005) where it has previously been recorded. According to records, the OTU assigned to the genus *Osphronemus* is likely the Giant Gourami (*Osphronemus gouramy*), introduced from mainland Southeast Asia for fisheries. It is known from Beratan (Sentosa *et al*. 2013), Rawa Pening (Goeltenboth and Kristyanto 1994), and Toba (Whitten and

Damanik, 2012), although this eDNA sampling only detected it from Riam Kanan (Tanjung *et al*. 2013), and Chenderoh (Hashim *et al*. 2012) (12S). The OTU assigned to the Serrasalmidae family detected from Lake Matano (12S) is likely to be *Colossoma macropomum*, the only species within this family recorded by (Herder *et al*. 2012). This OTU matched with 100% Query Cover and 100% Identity to both the Tambaqui (*Colossoma macropomum*) and the Pirapitinga (*Piaractus brachypomus*). This could therefore, actually be a hybrid 'cachamoto' of a cross of these two species, as has been created for aquaculture purposes and introduced to Indonesia (López and Anzoátegui, 2012).

Several non-native cichlid fish species introduced from fisheries were also detected from the 12S data. The *Amphilophus* species recorded from Lake Beratan is likely to be the previously recorded Midas Cichlid (*Amphilophus citrinellus*) (Sentosa *et al*. 2013), introduced from Costa Rica and Nicaragua. This species was also previously recoded from Lake Batur (Sentosa and Wijaya, 2012; Budiasa *et al*. 2018) and Rawa Pening (Dahruddin *et al*. 2016) although not detected from these data. The *Amphilophus* OTU only detected from Lake Matano is likely the hybrid 'flowerhorn' cichlid a man-made hybrid complex, allegedly composed of parental species of the neotropical cichlid genera *Cichlasoma*, *Amphilophus* and *Paraneetroplus* (Herder *et al*. 2012). This species is invasive within Lake Matano, spreading rapidly, and posing a threat to native biodiversity (Herder *et al*. 2012).

Of all lakes sampled in this study, the Peacock Bass (*Cichla ocellaris*) has only been recorded from the Chenderoh Reservoir (Hashim *et al*. 2012), which was also the only site from which this species was amplified (12S). This is an alien species from Latin America, introduced for game fishing. The OTU assigned to the hybrid *Paraneetroplus synspilus* x *Amphilophus citrinellus* is a fish created in China and Taiwan, named the Red Parrot Fish. *A. citrinellus* has been recorded from Rawa Pening (Dahruddin *et al*. 2016), and so this hybrid fish is likely either the true species, or also present in addition.

Various Tilapia fish within the Pseudocrenilabrinae Superfamily were detected from the 12S data across all lakes apart from Semayang and Melintang. These included eight different OTUs, three of which were only found from Laut Tawar. It is likely that these OTUs belong to the Nile Tilapia (*Oreochromis niloticus*), or the Mozambique Tilapia (*O. mossambicus*), or some aquaculture hybrids of these Tilapia species which are commonly introduced from Africa for aquaculture in Indonesia (Green *et al*. 1978; De Silva, 1987; Muchlisin *et al*. 2009; Muchlisin and Azizah, 2009; Wijopriono *et al*. 2010; Hashim *et al*. 2012; Sentosa and Wijaya, 2012; Muchlisin, 2012; Herder *et al*. 2012; Oktavia and Faoziyah, 2016; Mardiah *et al*. 2016; Dahruddin *et al*. 2016; Budiasa *et al*. 2018).

The Western Mosquitofish (*Gambusia affinis*) was only detected from one inlet transect from Laut Tawar from both the 12S and 16S data. *Gambusia. affinis* is an invasive species from North America, introduced to China (along with many other tropical countries) in the early 1900s for mosquito control (Eidman, 1989; Siriwardena, 2010). *G affinis* is an aggressive invasive species, associated with the decline or eradication of native fish populations, as well as other non-target insect species, particularly damselflies. The *Poecilia* OTU detected from Beratan (12S) is likely to belong to the Guppy, (*Poecilia reticulata*) (Green *et al*. 1978; Sentosa and Wijaya, 2012; Budiasa *et al.* 2018). This is also an invasive species introduced from Central America for mosquito control (Jordan, 2008). The Green Swordtail (*Xiphophorus hellerii*) was also detected (12S) from Beratan (Sentosa *et al*. 2013), Buyan (Dahruddin *et al*. 2016, Green *et al*. 1978) and Laut Tawar (Muchlisin *et al*. 2009; Muchlisin and Azizah, 2009; Muchlisin *et al*. 2010; Muchlisin, 2012) where it has previously been recorded. Similar to *P. reticulata* and *G. affinis*, *X. helleri* is an aggressive invasive thought to be introduced for mosquito control and later maintained as an ornamental fish (Maddern, 2009).

The problematic invasive Amazon Sailfin Catfish (*Pterygoplichthys pardalis*) was detected (12S) where it has previously been recorded from Laut Tawar (Muchlisin *et al*. 2009), but not where it was previously recorded from Lake Matano (Herder *et al*. 2012).

One Ambassidae OTU was detected from Lake Toba (12S) from all samples at very high read counts per sample (between 531,201 and 164,315 reads with an average of 326,433 reads after bioinformatic filtering and custom 0.5% background filter, but before read normalisation to 9,000 reads). This is highly likely to be the invasive alien discussed in the Introduction, the Glassfish (*Parambassis siamensis*) introduced to Lake Toba in 2013 (Kartamihardja *et al*. 2015). The unintentional introduction of this species caused a sharp decline in the local 'bilih fish' (*Mystacoleucus padangensis*) (Hedianto and Kartamihardja).

The Indo-Pacific, marine and freshwater species known as the Milkfish (*Chanos chanos*) was detected from Laut Tawar (12S) where it has not been previously recorded, although it is known from the local area of Aceh from the Pante Radja Canal, Aceh River and Cut River (Muchlisin *et al*. 2009). This species has been farmed in aquacultural ponds since the 1400s – 1600s (FAO, 2018).

The Grass carp (*Ctenopharyngodon idella*) was detected from Rawa Pening (16S) where it has previously been recorded (Dahruddin *et al*. 2016) although not where it has been previously recorded from Lake Laut Tawar (Muchlisin *et al*. 2009; Muchlisin and Azizah, 2009; Muchlisin, 2012), Lake Toba (Fishbase 2017a), the Chenderoh Reservoir (Kah-Wai

and Ali, 2000), Lake Batur (Kartamihardja, 2012), and Lake Beratan (Sentosa *et al*. 2013; Whitten *et al*. 1996).

5.5.5 Unexpected fish species

The Mackerel Scad (*Decapterus macarellus*) and Bigeye Scad (*Selar crumenophthalmus*) were detected (12S and COI) from Laut Tawar and Matano respectively. These scad species are usually marine based, but could have been detected from these lakes as pollution from cooking from nearby houses or local restaurants. The OTU assigned to Engraulidae (anchovy fish) from Lake Semayang (12S) could be one of five species recorded from Indonesia: *Coilia lindmani, Coilia borneensis, Lycothrissa crocodilus, Setipinna melanochir* or *Thryssa scratchleyi* (Fishbase 2017b). Usually anchovy type fish are marine, although there are some brackish and freshwater species. A tuna species within the *Auxis* genus was detected from Lake Matano. Tuna are strictly marine species which cannot survive in freshwater, and so this fish was likely a result of human waste pollution, as *Auxis* species such as *A. thazard* are native to the marine waters of the Malay Archipelago and eaten locally (Rivai *et al*. 2018).

5.5.6 Mammal species

There was one mammal OTU detected from the 12S data, also detected from the 16S data, assigned to the Miltred Leaf Monkey (*Presbytis melalophos*) with 100% Query Cover and Identity, found only from samples from the Chenderoh Reservoir. This monkey is an endangered species (IUCN, 2018a), found from the rainforests of Peninsular Malaysia, Borneo and Sumatra (Oates, *et al*. 1994), thought to be locally extinct (Davies and Oates, 1994). The mammal species detected from the 16S data were all native to Southeast Asia, or are the expected domestic species. The domestic dog (*Canis lupus familiaris*) and the domestic cat (*Felis catus*) were detected from a number of lakes, as is to be expected when domestic dwellings occur close to the water. The Crab-Eating Macaque (*Macaca fascicularis*) was detected from the Chenderoh Reservoir, a common species found in Malaysia (Ong and Richardson, 2008). Two agricultural species were detected, the Water Buffalo (*Bubalus bubalis*) from Laut Tawar, Toba and Chenderoh, and the domestic goat (*Capra hircus*) from the Chenderoh Reservoir.

There was an OTU assigned to the Feliformia family detected from the Chenderoh Reservoir samples, which matched with 100% Query Cover and 94% Identity to the Spotted Linsang, *Prionodon pardicolor*. This linsang species does not occur in Malaysia, although its close relative and the only other species within this genus - The Banded Linsang

Prionodon linsang - does. *Prionodon linsang* does not have a 16S gene or whole mitochondrial gene entry to NCBI, and so it is likely that this OTU belongs to this species. A phylogenetic tree (Figure 5.26) was created using 16S sequences from all Feliformia species extant from Malaysia (Mammals of Malaysia, 2018), which suggests that this OTU falls most closely amongst the *Prionodon* genus.

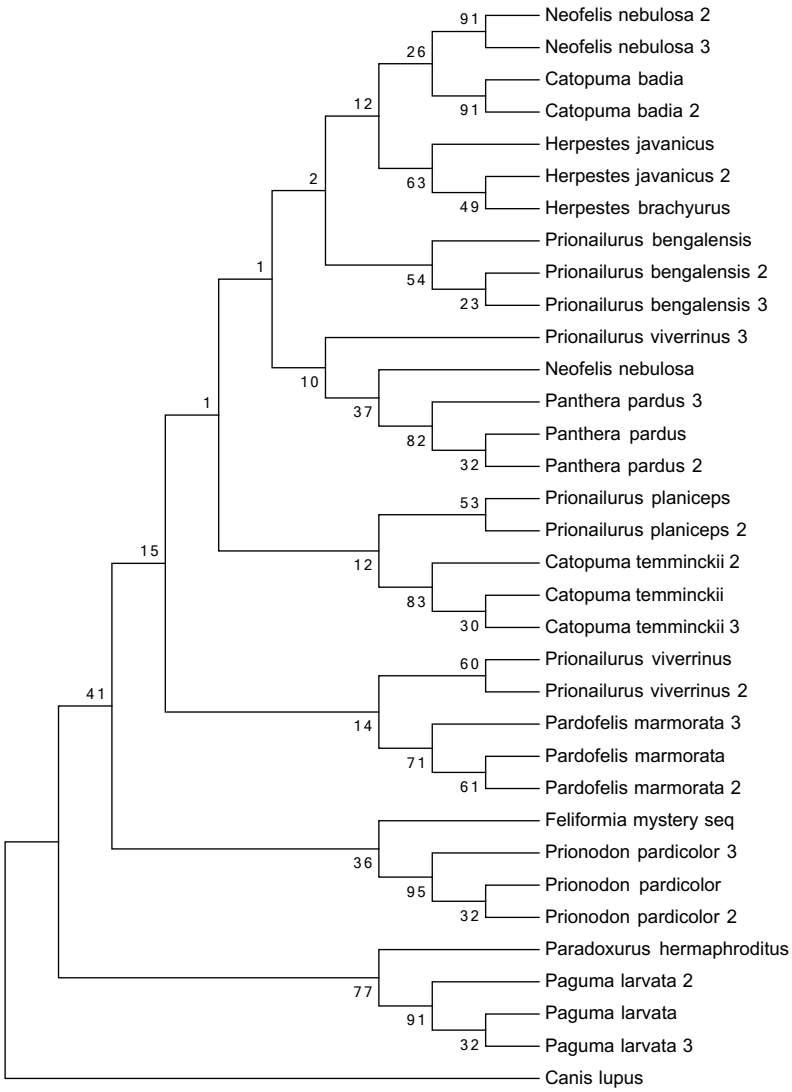


Figure 5.26. Phylogenetic tree of 16S mitochondrial gene regions of all Feliformia species from Malaysia. This is a neighbour-joining tree was created using the Maximum Likelihood method based on the Kimura 2-parameter model. Sequences were collected of the ~ 90 bp region of the 16S gene, using sequences from NCBI. The Feliformia mystery sequence falls within the *Prionodon* branch, however its position as sister to *Prionodon pardicolor* is not supported with a bootstrap value of <50%.

5.5.7 Other vertebrates

The 16S data also detected the Asian Common Toad (*Duttaphrynus melanostictus*) from Lake Toba and the Riam Kanan Reservoir, as well as the Asian Giant Toad (*Phrynoidis asper*) from the Chenderoh Reservoir. These two frog species are common, and widespread throughout Southeast Asia (Ngo and Ngo, 2013; IUCN, 2018b).

5.5.8 Challenges faced during this study

There were many challenges faced during this study relating to tropical field work, eDNA sampling, bioinformatic processing and interpreting results with respect to the available literature.

As the NCBI database is incomplete, and particularly lacking in species from Southeast Asian countries, many species assignments are not yet possible. A certain species may not exist in the database, or it may only have one gene sequenced which is not the target gene. For example, *Rasbora baliensis*, an endemic cyprinid fish to Bali, has six entries to NCBI, all of which are for COI. In this study, *Rasbora baliensis* was not detected from any of the Balinese lake samples. This could be due to 1) the absence of this fish in the lakes, 2) the absence of eDNA from this fish in the water sampled, 3) the absence of amplification of the target eDNA with the primers used, or 4) the absence of a voucher sequence in the database. As the COI primers used in this study preferentially amplified microfauna, meiofauna and microalgae, with very few Chordates amplified, it is impossible to know whether this result is a false negative or true negative. Although several *Rasbora* species were amplified using 12S and 16S markers, (at times with identities of 97-98%), as the NCBI database is lacking in vouchers for 12S and 16S for this species as well as many others, it would not be possible for a match to be found.

The dominant fish OTUs were within the cyprinid family, and many sequences were only possible to assign at this level. Cyprinidae is the largest fish family, with 210 genera and more than 2010 species, possibly making up around 20% of freshwater fishes, and 8% of all fishes, the greatest centre of diversity being China and Southeast Asia (Berra, 1977; Nelson, 1994, Berra, 2001). Other studies have had success in discriminating between cyprinids using eDNA metabarcoding using primers which target the cytochrome b gene. Keskin *et al*. (2016) identified 23 fish OTUs to species level from one lake in Turkey, 15 of which were within the Cyprinidae family. For studies in Southeast Asian freshwaters, based on metabarcoding with these primer pairs, it would be beneficial to use cyprinid specific metabarcoding primers.

For the COI data, there were 290,581 reads from 59 OTUs removed out of a total of 3,870,903 from 600 OTUs after initial bioinformatic filtering due to their low query cover (< 55), as well as 29,533 reads (4 OTUs) removed due to their presence in negative controls. Although low quality BLAST hits were sometimes consistent in their appearance from particular samples, it may be that they derive from chimeras or sequencing errors and so were removed. For example, OTU63, assigned to the wading bird, the Common Sandpiper (*Actitis macularia*), had a Query Cover of 30 and Identity of 84, but was only found in samples from Lake Toba (5/10 samples). The first 10 hits of OTU63 all assigned to *A. macularia*, although the following 10 hits assigned to *Canthophrys gongota*, a freshwater fish in the loach family. Similarly, OTU80, assigned to Bilateria (Query Cover 32, Identity 81) was only found in samples from Lake Tamblingan (19/19). Although these low-quality reads showed some consistency in their appearance in particular samples, they were removed from the analysis as their assignments cannot be trusted, and they could be a result of sequencing error or chimeras. There were only nine OTUs from the COI data which matched to their assignments with Query Cover of 100 and Identity of 100.

Some species hits from the BLAST search resulted in a Query Cover of 100 and Identity of 99 (accepted species level assignment) but were written with 'cf.' between the genus and species name e.g. *Thermocyclops* cf. *taihokuensis*. As this indicates that the taxonomic assignment of the specimen was unclear due to preservation issues, only a genus level (in this case *Thermocyclops*) was accepted.

It was not possible for the 12S marker used to distinguish between the Pseudocrenilabrinae subfamily of African and Middle Eastern cichlid fish, as BLAST results returned 100% identity and query cover for many different species within this subfamily.

Synonyms of some fish species made it difficult to assess the previously recorded local ichthyofaunal biodiversity. For example, the Silver Barb (also referred to as the Java Carp or Java Barb), *Barbonymus gonionotus* (the currently accepted name by fishbase.com and IUCN) has nine different synonyms listed on fishbase.com. This species is sometimes referred to as '*Barbodes gonionotus*' (Kah-Wai and Ali, 2000; Wijopriono *et al*. 2010; Kurniawan and Subehi, 2016), but was originally named *Puntius gonionotus* (Bleeker, 1849), and later *Puntius javanicus* (Bleeker, 1855). Green *et al*. (1978) refer to this species as *Barbus gonionotus*, whilst Hutarabat *et al*. (1986) use *Puntius javanicus*.

5.5.9 Interpretation of the results

The aquatic eDNA sampling approach used here to monitor the freshwater biodiversity of lakes in the Malay Archipelago was successful in amplifying a range of vertebrates and invertebrate species. The 12S and 16S markers were most useful in identifying vertebrates, whilst the COI marker mostly amplified microfauna, meiofauna and microalgae. It is unlikely however that the sampling approach used was sufficient to detect all fish species present at the time of sampling, as many more species were previously recorded from the literature per lake.

The factors affecting OTU richness for the 12S and 16S markers which mostly amplified vertebrate species were productivity (higher richness in eutrophic or mesotrophic lakes), depth (higher richness in shallow lakes), area (higher richness in small and medium sized lakes) and altitude (higher richness in lowland lakes). It is expected that eutrophic lakes which have a higher trophic productivity would have a higher richness. It is also expected that more shallow lakes would have a higher richness, as discussed in Chapter 4, most biodiversity is found in the light filled shallow zones of the lake. It was however not expected that smaller lakes had a higher OTU richness, as lake species richness is determined by size. This pattern may have been observed due to the deeper lakes also being oligotrophic. Lowland lakes may have had a higher OTU richness due to the increased influx of eDNA from rivers, which do not generally enter high altitude lakes, especially isolated volcanic calderas.

Some lakes which have higher OTU richness may be due to their influx of DNA from rivers, e.g. Semayang / Melintang / Chenderoh, compared to isolated meromictic lakes e.g. Toba / Matano / Batur. Reservoirs with rich ichthyofaunal diversity are thought to be incapable of sustaining high fish yields, even in the presence of lacustrine or lacustrine-adapted fish species (Amarasinghe and De Silva, 2015), which may be the case with regards to the Chenderoh Reservoir, which showed a high ichthyofaunal diversity. The Chenderoh Reservoir had consistently higher OTU richness than the other lakes sampled. This could be a real pattern observed, possibly due to the presence of the Perak River flowing through this reservoir, increasing the fish biodiversity present, or it could also be to do with the success of the preservation of the filters at the time of collection. The samples from the Chenderoh Reservoir were the only ones to be filtered and immediately shipped on dry ice to Denmark where they were extracted at the GeoGenetics laboratory at the Natural History Museum of Denmark. Other samples were either extracted in Indonesia at the IBRC laboratory and then shipped to Bangor University, then to Copenhagen University, or, they were filled with an

EDTA buffer and shipped to the Natural History Museum of Denmark (samples from Lake Riam Kanan and the removed Lake Singkarak).

5.5.10 Possible improvements to this study

For the sampling strategy, a transect approach using a boat was implemented to allow rapid sampling of the maximum area possible given the time, equipment, and ability to access the lake habitat. However, it would have been most effective in terms of capturing total biodiversity to sample each lake by sampling at regular points across the entire surface, at depth, and also at more points around the edge of the lake.

If more lakes were sampled then patterns related to area / depth / productivity could be better understood. Other factors were measured which were not included in the analysis (pH, temperature, turbidity, dissolved oxygen, lake depth).

Nine OTUs assigned to the Cyprinidae family were detected from Beratan, Laut Tawar, Semayang and Melintang. It was not possible to assign these OTUs to a lower taxonomic rank, indicating the need for more barcoding work within the Cyprinidae family to be done in Indonesia and Malaysia. The addition of a universal cyprinid primer would have also been beneficial for this region where cyprinid diversity is particularly high. The 12S primer which targets Teleost fish did not discriminate well between cyprinid fish OTUs, and so was likely too broad for this highly diverse region.

Although the bioinformatic filtering of sequences according to their presence in at least 2/3 PCRs will limit the number of false positives, it is likely that this technique does create false negatives. After filtering, some OTUs were lost which had been assigned to genus or species level. For example, when only filtering for a minimum of 2 copies in 2/3 PCR replicates, there were several OTUs from the 12S dataset which were removed. An OTU from the Tasik Chenderoh data, assigned to the *Devario* genus, with 100% query cover and 98% match to the Bengal Danio (*Devario devario*) was removed as this OTU contained only 55 reads. This OTU is likely to actually belong to the Queen Danio (*Devario regina*) known from the area in Perak (Ikhwanuddin *et al*. 2017), but not found in the NCBI database. There was also an OTU assigned to the Sumatran River Sprat (*Clupeichthys goniognathus*) from the same lake with only 45 reads in total from two samples, and although this OTU matched with 100% query cover and 100% identity, it is likely to be a wrongly identified Perak River Sprat (*Clupeichthys perakensis*) sequence, the fish that is native to Perak, and a species which matched to many thousands of reads from this dataset. Also removed after filtering were six OTUs all assigned to Fuentesi's Wrasse (*Pseudolabrus fuentesi*) with 55

reads in total, and 4-9 reads per sample, likely a result of background contamination from the positive controls. An OTU assigned to the Javanese Rice Fish (*Oryzias javanicus*) with 100% query cover and 100% identity, composed of 28 reads across four samples, all of which were negative controls, was also removed. Another OTU only found in one negative control was assigned to the Asian Common Toad (*Duttaphrynus melanostictus*), which is native to Indonesia and Malaysia. Similarly, an OTU assigned to the Asian Water Monitor (*Varanus salvator*) with 100% query cover and 100% identity, only found in one negative with only six reads was removed. The removal of these OTUs may, in some cases, be creating false negatives. However, a filtering system must be implemented which removes false positives, and it is clear from comparing the OTU tables with either filtering for a minimum of two copies, or filtering for a minimum of 20 copies, that removing OTUs with low read abundance helps to remove the low read abundance 'tails' of OTUs clearly assigned to one particular taxonomic level. It is unclear how these rare sequences appeared in the sample, whether through lab contamination or aerial contamination whilst sampling. It is interesting however that some of these rare OTUs assigned to local species only appeared in some sample negatives, as opposed to many eDNA samples with some spill over into negatives. These examples demonstrate that it is of uttermost importance to sequence negative controls to understand where reads are occurring and not overestimate what diversity is present in the data.

The only accepted detection for the Climbing Perch (*Anabas testudineus*) was from Semayang and Melintang, although sequences were filtered from Lake Beratan, Riam Kanan, Laut Tawar which may have been cross contamination from the positive control of the same species. Therefore, to improve studies such as this, a non-native species should be used for the positive control, or if native species are used, they should be incorporated into a mock community of specific low concentrations. A similar pattern was observed by Hanfling *et al*. (2016), who also suggest this diluted mock community or different target species as negative controls as a possible solution.

5.5.11 Suggestions for future eDNA research in the tropics

At lakes Semayang and Melintang, the high level of turbidity observed made processing the water samples through the 0.22 µm filters too difficult to allow a total volume of 500 ml to be processed. At these sites therefore, only 100 ml could be filtered per Sterivex filter. For future eDNA studies in the tropics, it may be beneficial to first use a wider pore filter

followed by a fine pore filter, or use additional filters to allow a larger total volume to be processed.

Although barcoding efforts (particularly of ichthyodiversity in Indonesia) are ongoing, the COI barcode region continues to be the focus of barcoding attempts rather than other mitochondrial regions which could be more suitable (such as 12S). If whole mitogenome sequencing of biodiversity could be used in barcoding studies of this region, this would provide greater specificity in public databases from which to compare eDNA metabarcoding data (e.g. Dahruddin *et al*. 2016).

5.5.12 *Implications of aquatic eDNA monitoring in Southeast Asia*

Indonesian researchers note that lakes must be restored and protected to enhance their ecosystem services, particularly those linked to other aquatic ecosystems (Haryani, 2016). The use of eDNA metabarcoding, as demonstrated here, can provide a large amount of taxonomic information from few samples collected within a short period of time. This study used few samples across a small number of sample points. However, if a specific area of interest were to be more intensively monitored (e.g. the Danau Sentarum National Park) over many sampling occasions, this could provide extensive biodiversity data as a baseline from which to then monitor changes over time as a result of either conservation protection, or anthropogenic impact from the threats of hydrological dams, for example.

This study has shown the potential for eDNA metabarcoding in monitoring the distribution of invasive species, which are evidently a problem in Southeast Asian freshwaters. More studies are needed to understand the role of exotics in the geographical variability in lake and reservoir fish yields. Based on these data generated from eDNA sampling of the lakes of the Malay Archipelago, invasive species are dominating lacustrine environments, possibly at the cost of the exclusion of rare native species, or of native fish which are significant for local fisheries. Another interesting and important avenue for future research in fisheries and biodiversity conservation could be to assess the relationship between stocking density and CBF fish yield, also with respect to rare species. It is thought that there are density-dependent factors in force that create optimum levels of CBF production (Amarasinghe and De Silva, 2015), something which could be monitored using eDNA metabarcoding, particularly if relative read abundance can be used as a rough measure of density.

Indigenous cyprinid species in Asia which occupy lower trophic levels play a significant role in reservoir and lake trophic dynamics, and can withstand exploitation due to

their high turnover rates. The stocking of commonly exploited African cichlid species mostly leads to incomplete exploitation of predominant fishery sources (Piet and Vijverberg, 1998). The use of eDNA metabarcoding could therefore also be useful in monitoring the response of introduced fisheries species according to the levels of pre-existing ichthyofaunal diversity, particularly of cyprinids.

Malaysia is one of the more affluent countries in Southeast Asia, and consequently has better infrastructure than some other Southeast Asian countries. One asset of which, is the permanent employment of an officer responsible for compiling statistics of aquaculture and inland capture fisheries by the State Department of Fisheries in each district (Coates, 2002). Malaysia also has the largest number of technical persons trained per country under the International Network on Genetics in Aquaculture (INGA), with 38 persons compared to just 6 in Indonesia (De Silva, 2010). It is possible therefore that in Malaysia, the use of eDNA sampling could be employed to monitor freshwater biodiversity through government branches, as is beginning in Europe.

This study was conducted in the summer, in the dry season. Physiological measurements however vary between wet and dry season - for example, temperatures at Danau Batur range from 22 – 25 ºC, pH 7.11 – 8.82 in the rainy season and 8.55 – 8.61 in the dry season, DO 6.43 – 7.7 in the rainy season and 7.2 – 9.3 in the dry season, Turbidity is 3.39 – 5.13 NTU in the rainy season 2.4 – 3.7 NTU in the dry season (Suryaningtyas and Ulinuha, 2016). It would therefore be interesting for a more intensive sampling strategy to be implemented to allow temporal analysis with respect to these variables.

Future studies in Indonesia using eDNA could benefit from a more targeted approach to explore specific local hypotheses relating to comparable areas of pollution or anthropogenic impact. For example, in East Kalimantan, the Mahakam connected lakes measured in this study (Melintang and Semayang) have nearby lakes which do not appear to be as impacted by runoff from mining and logging. Using eDNA to assess microbial, invertebrate and vertebrate diversity between these sites could help to illuminate the effects of the rampant mining industry on biodiversity in this region.

Another possible future avenue for research is that relating to lake stratification. Future stratification of Lake Batur, Matano and Toba is likely to be caused by climate change, when increasing temperature and evaporation will shift the thermocline layers, although this is not fully understood (Haryani, 2016).

5.6 Conclusions

This study highlights the success of eDNA metabarcoding in monitoring the biodiversity of Southeast Asia, particularly of ichthyofaunal species for the first time. Thousands of eDNA reads were successfully amplified and assigned to many native, invasive and rare species of conservation concern. Although there are improvements to be made on this sampling strategy, it was overall successful in detecting some of the known biodiversity from this mega-diverse region. Although patterns of OTU richness and community composition with regards to habitat variables were observed, more sampling at both the temporal and spatial scale as well as increasing the number of sites would help to further understand the role of these features in driving patterns of local biodiversity.

5.7 References

Abell, R., Thieme, M. L., Revenga, C. *et al*. 2008. Freshwater ecoregions of the world: a new map of biogeographic units for freshwater biodiversity conservation. *BioScience*, 58, 403–414.

Abery, N.W., Sukadi, F., Budhiman, A.A., Kartamihardja, E.S., Koeshendrajana, S. and De Silva, S.S., 2005. Fisheries and cage culture of three reservoirs in west Java, Indonesia; a case study of ambitious development and resulting interactions. *Fisheries Management and Ecology*, 12(5), pp.315-330.

Ali, A.B. 1996. Chenderoh Reservoir, Malaysia: The conservation and wise use of fish biodiversity in a small flow-through tropical reservoir. *Lakes and Reservoirs: Research and Management*, 2:, 17–30.

Amarasinghe, U.S. and De Silva, S.S., 2015. Fishes and fisheries of Asian inland lacustrine waters. In Freshwater Fisheries Ecology (pp. 384-403). John Wiley & Sons, Ltd.

Amarasinghe, U.S. and Welcomme, R.L., 2002. An analysis of fish species richness in natural lakes. *Environmental Biology of Fishes*, 65(3), pp.327-339.

Argillier, C., Caussé, S., Gevrey, M., *et al*. 2013. Development of a fish-based index to assess the eutrophication status of European lakes. *Hydrobiologia* 704, 193-211.

Arifin, A. *et al*., 2015. Keberadaan ikan hias eksotik di Danau Batur dan Beratan, Bali. *Prosiding simposium nasional ikan hias.* The existence of exotic ornamental fish in Batur Lake and weight, Bali. *Proceedings of the national symposium on ornamental fish*.

Arthana, I. W. THE CHARACTERISTIC OF WATER QUALITY AT BATUR LAKE, KINTAMANI DISTRIC, BALI PROVINCE. Jurnal Bumi Lestari, Volume 11 No. 1, Pebruari 2011, hlm. 40-49

Aylagas, E. *et al*., 2016. Benchmarking DNA Metabarcoding for Biodiversity-Based Monitoring and Assessment. *Frontiers in Marine Science*, 3 (November).

Bálint, M., Nowak, C., Márton, O., Pauls, S., Wittwer, C., Aramayo, J.L., Schulze, A., Chambert, T., Cocchiararo, B. and Jansen, M., 2017. Twenty-five species of frogs in a litre of water: eDNA survey for exploring tropical frog diversity. *bioRxiv*, p.176065.

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.

Bemmelen, R.W., 1970. *The geology of Indonesia* (Vol. 1). Martinus Nijhoff.

Berra, T. M., and Berra, R. M. 1977. A temporal and geographical analysis of new teleost names proposed at 25 year intervals from 1869-1970. *Copeia* 1977(4), 640-647.

Berra, T.M., 2001. *Freshwater fish distribution*. Academic Press.

Bolotov, I.N., Bespalaya, Y.V., Gofarov, M.Y., Kondakov, A.V., Konopleva, E.S. and Vikhrev, I.V., 2016. Spreading of the Chinese pond mussel, Sinanodonta woodiana, across Wallacea: One or more lineages invade tropical islands and Europe. *Biochemical systematics and ecology*, 67, pp.58-64.

Bonar, S. A., Hubert, W. A., Willis, D. W. (eds). 2009. *Standard Methods for Sampling North American Freshwater Fishes*. American Fisheries Society, Bethesda, Maryland.

Brander, K., M. 2007. Global fish production and climate change. *Proceedings of the National Academy of Sciences of the USA*, 104, 19709–19714.Brooks, J.L. 1950. Speciation in ancient lakes. The Quarterly Review of Biology 25:30–60, http://dx.doi.org/10.1086/397375

Budiasa, I. W., Santosa, I. G. N., Ambarawati, I. G. A. A., Suada, I. K., Sunarta, I. N., Shchegolkova, N. 2018. Feasibility study and carrying capacity of Lake Batur ecosystem to preserve tilapia fish farming in Bali, Indonesia. BIODIVERSITAS. ISSN: 1412-033X Volume 19, Number 2, March 2018 E-ISSN: 2085-4722 Pages: 613-620

Cerwenka, A. F., Wedekind, J. D., Hadiaty, R. K., Schliewen, U. K., Herder F. 2012. Alternative egg-feeding tactics in *Telmatherina sarasinorum*, a trophic specialist of Lake Matano's evolving sailfin silversides fish radiation. *Hydrobiologia* 693: 131–139, http://dx.doi.org/10.1007/s10750-012-1099-8

Cilleros, K., Valentini, A., Allard, L., Dejean, T., Etienne, R., Grenouillet, G., Iribar, A., Taberlet, P., Vigouroux, R. and Brosse, S., 2018. Unlocking biodiversity and conservation studies in high diversity environments using environmental DNA (eDNA): a test with Guianese freshwater fishes. *Molecular ecology resources*.

Civade, R., Dejean, T., Valentini, A., Roset, N., Raymond, J.C., Bonin, A., Taberlet, P. and Pont, D., 2016. Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. *PloS one*, 11(6), p.e0157366.

Coates, D. 1985. Fish yield estimates for the Sepik River, Papua New Guinea, a large floodplain system east of "Wallace's Line". *Journal of Fish Biology* 27: 431-443.

Coates, D. 2002. Inland capture fishery statistics of Southeast Asia: current status and information needs. *RAP publication*, 11, p.114.

Collen, B., Whitton, F., Dyer, E.E., Baillie, J.E., Cumberlidge, N., Darwall, W.R., Pollock, C., Richman, N.I., Soulsby, A.M. and Böhm, M., 2014. Global patterns of freshwater species diversity, threat and endemism. *Global Ecology and Biogeography*, 23(1), pp.40-51.

Creer, S. *et al*., 2016. The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution.*

Crowe, S. A., O'Neill, A. H., Katsev, S., Hehanussa, P., Haffner, G. D., Sundby, B., Mucci, A., Fowle, D. A. 2008. The biogeochemistry of tropical lakes: A case study from Lake Matano, Indonesia. *Limnology and Oceanography* 53: 319–331, http://dx.doi.org/10.4319/lo.2008.53.1.0319

Dahlen, B.F. 1993. Hydropower in Malaysia. Tenaga Nasional Berhad (TNB), Malaysia, 184 p.

Dahruddin, H., Hutama, A., Busson, F., Sauri, S., Hanner, R., Keith, P., Hadiaty, R. and Hubert, N., 2016. Revisiting the ichthyodiversity of Java and Bali through DNA barcodes: taxonomic coverage, identification accuracy, cryptic diversity and identification of exotic species. *Molecular Ecology Resources*, 17(2), pp.288-299.

Davies, G. and Oates, J. eds., 1994. *Colobine monkeys: their ecology, behaviour and evolution*. Cambridge University Press.

De Silva, S.S. and Davy, F.B., 2010. Success stories in Asian aquaculture. IDRC, Ottawa, ON, CA.

Djajadireja, R., S. Fatimah and Z. Arifin, 1977. Jenis-Jenis Ikan Ekonomis Penting. Ditjen Perikanan, Deptan, Jakarta. Important Economic Fish Types. *Directorate General of Fisheries, Ministry of Agriculture, Jakarta.*

Dudgeon, D. 2010. Prospects for sustaining freshwater biodiversity in the 21st century: linking ecosystem structure and function. *Current Opinion in Environmental Sustainability*, 2, 422–430.

Eidman, H.M., 1989. Exotic aquatic species introduction into Indonesia. Exotic aquatic organisms in Asia. *Asian Fisheries Society Special Publication*, 3, pp.57-62.

European Communities. 2000. Directive 2000/60/EC, Establishing a framework for community action in the field of water policy. *Official Journal of the European Communities* L 327, 1-71.

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. and Pfrender, M.E., 2017. Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(9), pp.1362-1374.

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. and Lodge, D.M., 2016. Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular ecology resources*, 16(1), pp.29-41.

Foster, Z.S.L. *et al*., 2017. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data T. Poisot, ed. *PLOS Computational Biology*, 13(2), p.e1005404.

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(2).

Giller, P. S., Hillebrand, H., Berninger, U. G. *et al*. 2004. Biodiversity effects on ecosystem functioning: emerging issues and their experimental test in aquatic environments. *Oikos*, 104, 423–436.

Goeltenboth, F. and Kristyanto, A.I.A., 1994. Fisheries in the Rawa pening reservoir, Java, Indonesia. *Internationale Revue der gesamten Hydrobiologie und Hydrographie*, 79(1), pp.113-129.

Gomes, G.B., Hutson, K.S., Domingos, J.A., Chung, C., Hayward, S., Miller, T.L. and Jerry, D.R. 2017. Use of environmental DNA (eDNA) and water quality data to predict protozoan parasites outbreaks in fish farms. Aquaculture, 479, pp.467-473.

Gray, S. M., Dill, L. M., McKinnon, J. S. 2007. Cuckoldry incites cannibalism: male fish turn to cannibalism when perceived Malili Lakes alien fishes 533 certainty of paternity decreases. *The American Naturalist*. 169: 258–263, http://dx.doi.org/10.1086/510604

Gray, S. M., Dill, L. M., Tantu, F. Y., Loew, E. R., Herder, F., McKinnon, J. S. 2008b. Environment-contingent sexual selection in a colour polymorphic fish. Proceedings of the Royal Society London B 275: 1785–1791, http://dx.doi.org/10.1098/rspb.2008.0283

Gray, S. M., McKinnon, J. S., Tantu, F. Y., Dill, L. M. 2008a. Sneaky egg-eating in *Telmatherina sarasinorum*, an endemic fish from Sulawesi. Journal of Fish Biology 73: 728–731, http://dx.doi.org/10.1111/j.1095-8649.2008.01949.x

Green, J., Corbet, S.A., Watts, E. and Lan, O.B., 1978. Ecological studies on Indonesian lakes. The montane lakes of Bali. *Journal of Zoology*, 186(1), pp.15-38.

Hänfling, B., Lawson Handley, L., Read, D.S., Hahn, C., Li, J., Nichols, P., Blackman, R.C., Oliver, A. and Winfield, I.J., 2016. Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular ecology*, 25(13), pp.3101-3119.

Hardjamulia, A. and Suwignyo, P., 1988. Present status of the reservoir fishery in Indonesia. In *Reservoir fishery management and development in Asia: proceedings of a workshop held in Kathmandu, Nepal, 23-28 Nov. 1987*. IDRC, Ottawa, ON, CA.

Haryani, G. S. 2016. LAKE RESTORATION IN INDONESIA: A RISK BASED ECOHYDROLOGY APPROACH. Proceedings of the 16th World Lake Conference.

Haryono. 2016. Iktiofauna di Danau Semayang-Melintang kawasan Mahakam Tengah, Kalimantan Timur. *Jurnal Iktiologi Indonesia*. Volume 6, Nomor 1, Juni 2006. Ichthyofauna at Semayang-Melintang Lake in the Middle Mahakam region, East Kalimantan. *Indonesian Ichthyology Journal. Volume 6, 1st June 2006.*

Hashim, Z.H., Zainuddin, R.Y., Shah, A.S.R.M., Sah, S.A.M., Mohammad, M.S. and Mansor, M., 2012. Fish checklist of Perak River, Malaysia. *Check List*, 8(3), pp.408-413.

Hedianto, D.A., and Kartamihardja, E. S. 2014. Karakteristik biologi dan dampak introduksi ikan kaca (*Parambassis siamensis*, Fowler 1937) di Danau Toba. Biological characteristics and the impact of glass fish introduction (*Parambassis siamensis*, Fowler 1937) in Lake Toba. pp.139–152.

Herder, F., Nolte, A., Pfaender, J., Schwarzer, J., Hadiaty, R. K., Schliewen, U. K. 2006. Adaptive radiation and hybridization in Wallace's Dreamponds: evidence from sailfin silversides in the Malili Lakes of Sulawesi. Proceedings of the Royal Society London B, 275: 2178–2195, http://dx.doi.org/10.1111/j.1558-5646.2008.00447.x

Herder, F., Pfaender, J., Schliewen, U. K. 2008. Adaptive sympatric speciation of polychromatic "roundfin" sailfin silverside fish in Lake Matano (Sulawesi). *Evolution.* 62: 2178–2195

Herder, F., Schliewen, U.K., Geiger, M.F., Hadiaty, R.K., Gray, S.M., McKinnon, J.S., Walter, R.P. and Pfaender, J., 2012. Alien invasion in Wallace's Dreamponds: records of the hybridogenic" flowerhorn" cichlid in Lake Matano, with an annotated checklist of fish species introduced to the Malili Lakes system in Sulawesi. *Aquatic Invasions*, 7(4).

Herre, A.W.C.T., 1940. New species of fishes from the Malay Peninsula and Borneo. *Bulletin of the Raffles Museum*, 16, pp.5-26.

Hidayaturrahmah, M. Kematian ikan nila pada budi daya keramba jaring apung di Desa Aranio dan Tiwingan Lama Kabupaten Banjar, Kalimantan Selatan. PROS SEM NAS MASY BIODIV INDON.

Volume 3, Nomor 1, Februari 2017 ISSN: 2407-8050. Halaman: 28-32 The death of nile tilapia on the floating net cage farming in Aranio and Tiwingan Lama Lama villages, Banjar District, South Kalimantan. *Proceedings of the National Seminar on the Indonesian Biodiversity Community.*

Holmlund, C.M. and Hammer, M., 1999. Ecosystem services generated by fish populations. *Ecological economics*, 29(2), pp.253-268.

Hubert, N. *et al*., 2015. DNA Barcoding Indonesian freshwater fishes: challenges and prospects. *DNA Barcodes*, 3(1), pp.144–169.

Hutarabat, J., Syarani, L. and Smith, M.A.K., 1986. Use of freshwater hyacinth Eichhornia crassipes in cage culture in Lake Rawa Pening, Central Java. In 1. *Asian Fisheries Forum, Manila (Philippines),* 26-31 May 1986.

Ikhwanuddin, M.E.M., Amal, M.N.A., Aziz, A., Sepet, J., Talib, A., Ismail, M.F. and Hashim, N.R., 2017. Inventory of fishes in the upper Pelus River (Perak river basin, Perak, Malaysia). Check List, 13(4), pp.315-325.

Irawan, E. 2016. THE GOVERNANCE OF LAKE RAWAPENING: AN INTERORGANIZATIONAL NETWORK ANALYSIS. *Proceedings of the 16th World Lake Conference*.

Ishak, H. O., Kusairi M. N. N. M., Raja A., and Kuperan, K. 1992. Malaysian Fisheries Policy - Search for New Grounds. *Marine Policy*. 16(6), 438-450, November.

Janosik, A.M. & Johnston, C.E., 2015. Environmental DNA as an effective tool for detection of imperiled fishes. *Environmental Biology of Fishes*, 98(8).

Kah-Wai, K. and Ali, A.B., 2000, February. Chenderoh Reservoir, Malaysia: Fish community and artisanal fishery of a small mesotrophic tropical reservoir. In *ACIAR PROCEEDINGS* (pp. 167-178). ACIAR; 1998.

Kartamihardja, E.S., 2012. STOCK ENHANCEMENT IN INDONESIAN LAKE AND RESERVOIRS FISHERIES. *Indonesian Fisheries Research Journal*, 18(2), pp.91-100.

Kartamihardja, E. S. 2015. Fish Stock Enhancement and Restocking of the Inland Waters of Indonesia: Lessons Learned. *Fish for the People.* Volume 13 Number 3: 2015.

Kartamihardja, E.S., Hedianto, D.A. and Umar, C., 2015. Strategi Pemulihan Sumber Daya Ikan Bilih (*Mystacoleucus padangensis*) Dan Pengendalian Ikan Kaca (*Parambassis siamensis*) Di Danau Toba, Sumatera Utara. *Jurnal Kebijakan Perikanan*. Selective Fish Recovery Strategy (*Mystacoleucus padangensis*) and Glass Fish Control (*Parambassis siamensis*) in Lake Toba, North Sumatra. Journal of Fisheries Policy. Indonesia, 7(2), pp.63-69.

Keskin, E., Unal, E.M. and Atar, H.H., 2016. Detection of rare and invasive freshwater fish species using eDNA pyrosequencing: Lake Iznik ichthyofauna revised. *Biochemical systematics and ecology*, 67, pp.29-36.

Kottelat, M., A.J. Whitten, S.N. Kartikasari & S. Wirjoatmodjo. 1993. *Freshwater Fishes of Western Indonesia and Sulawesi (Ikan Air Tawar Indonesia Bagian Barat dan Sulawesi)*. Periplus Editions Ltd. Indonesia. 293 p.

Kubečka, J., Hohausová, E., Matěna, J., *et al*. 2009. The true picture of a lake or reservoir fish stock: A review of needs and progress. *Fisheries Research* 96, 1-5.

Kurniawan, R. and Subehi, L., 2016. Aquatic Macrophytes and Fish Diversity of Various Tropical Lakes at the Main Islands in Indonesia. *Aquatic Biodiversity Conservation and Ecosystem Services* (pp. 3-12). Springer, Singapore.

Lehmusluoto, P. and Machbub, B. 1997. *National Inventory of the Major Lakes and Reservoirs in Indonesia.* Expedition Indodanau Technical Report.

Lim, N.K.M. *et al*., 2016. Next-generation freshwater bioassessment: eDNA metabarcoding with a conserved metazoan primer reveals species-rich and reservoir-specific communities. *Royal Society Open Science*, 3(11).

Lodge, D.M., Turner, C.R., Jerde, C.L., Barnes, M.A., Chadderton, L., Egan, S.P., Feder, J.L., Mahon, A.R. and Pfrender, M.E., 2012. Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology*, 21(11), pp.2555-2558.

López, P. and Anzoátegui, D., 2012. Crecimiento del hibrido Cachamoto (Colossoma Macropomum x Piaractus Brachypomus) en un sistema de recirculación de agua. Growth of the Cachamoto hybrid (*Colossoma Macropomum x Piaractus Brachypomus*) in a water recirculation system. Zootecnia Tropical, 30(4), pp.351-360.

Lumbantobing, D. N. Four New Species of the *Rasbora trifasciata* - Group (Teleostei: Cyprinidae) from Northwestern Sumatra, Indonesia. *Copeia* 4: 644–670

Mackenzie, D. I., Royle, J., A. 2005. Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*, 42, 1105–1114.

MacKinnon, K., 1996. The ecology of Kalimantan (Vol. 3). *Oxford University Press*.

Manuaba, I.P., 2007. Cemaran pestisida klor-organik dalam air Danau Buyan buleleng Bali. Jurnal kimia Juli 2007: 39-46. Characteristics of chlor-organic pesticides in Buyan Lake water Bali. *Journal of Chemistry*.

Mardiah, A. A., and Syandri, H. 2016. Fish diversity of the Singkarak Lake, Indonesia: present status and conservation needs. *Proceedings of the 16th World Lake Conference.*

Ministry of Environment Republic of Indonesia. 2012. Grand Design for Save Indonesian Lake Ecosystem. Ministry of Environment, RI. Jakarta. 72 p.

Muchlisin, Z.A. and Azizah, S., 2009. Diversity and distribution of freshwater fishes in Aceh waters, northern Sumatra Indonesia. *International Journal of Zoological Research*, 5(2), pp.62-79.

Muchlisin, Z.A., 2011. First report on introduced freshwater fishes in the waters of Aceh, Indonesia. *Indonesia – Arch. Pol. Fish*, 20, pp.129–135.

Muchlisin, Z.A., 2013. Morphometric Variations of Rasbora Group (Pisces: Cyprinidae) in Lake Laut Tawar, Aceh Province, Indonesia, Based on Truss Character Analysis. *Hayati Journal of Biosciences*, 20(3), pp.138-143.

Muchlisin, Z.A., Musman, M. & Siti Azizah, M.N., 2010. Length-weight relationships and condition factors of two threatened fishes, *Rasbora tawarensis* and *Poropuntius tawarensis*, endemic to Lake Laut Tawar, Aceh Province, Indonesia. *Journal of Applied Ichthyology*, 26(6), pp.949–953.

Muchlisin, Z.A., Thomy, Z., Fadli, N., Sarong, M.A. and Siti-Azizah, M.N., 2013. DNA barcoding of freshwater fishes from Lake Laut Tawar, Aceh Province, Indonesia. *Acta ichthyologica et piscatoria*, 43(1).

Murphy, B. R., Willis, D. W. (eds). 1996. *Fisheries Techniques, 2nd edn*. American Fisheries Society, Bethesda, Maryland.

Nasution, S. H. 2016. BIODIVERSITY AND CONSERVATION OF ENDEMIC FISH SPECIES IN SOME LAKES OF SULAWESI. Proceedings of the 16th World Lake Conference

Nasution, S.H. (2006). Pangkilang (Telmatherinidae) Ornamental Fish: An Economic Alternative for People Around Lake Towuti. *Proceedings International Symposium on The Ecology and Limnology of the Malili Lakes on March 20-22, 2006 in Bogor – Indonesia*. p 39-46.

Negara, I. K. W. and Susilo, M. D. 2015. Strategi pengembangan budidaya lele dumbo Clarias sp. Melalui program pengembangan usaha mina pedesaan perikanan budidaya di kabupaten buleleng. J. Manusia dan lingkungan. Strategy for development of Clarias sp. Through business development program for fishing rural farming in buleleng district. J Human and Environment**.** Vol. 22, No.3, November 2015: 365-371

Nelson, J.S., 1994. *Fishes of the World*. 3rd Edn., John Wiley and Sons, Inc., New York.

Ng, T.H., Tan, S.K., Wong, W.H., Meier, R., Chan, S.Y., Tan, H.H. and Yeo, D.C., 2016. Molluscs for sale: assessment of freshwater gastropods and bivalves in the ornamental pet trade. *PloS one*, 11(8), p.e0161130.

Ngo, B.V. and Ngo, C.D., 2013. Reproductive activity and advertisement calls of the Asian common toad Duttaphrynus melanostictus (Amphibia, Anura, Bufonidae) from Bach Ma National Park, Vietnam. *Zoological Studies*, 52(1), p.12.

Ninkovich, D., Shackleton, N.J., Abdel-Monem, A.A., Obradovich, J.D. and Izett, G., 1978. K–Ar age of the late Pleistocene eruption of Toba, north Sumatra. *Nature*, 276(5688), p.574.

Nomosatryo, Sulung, Cynthia Henny, Carry Ayne Jones, Celine Michiels and Sean A. Crowe. 2014. Karakteristik dan klasifikasi trofik di Danau Matano dan Danau Towuti Sulawesi Selatan. Prosiding Pertemuan Ilmiah Tahunan MLI. *Masyarakat Limnologi Indonesia.* Trophic characteristics and classification in Lake Matano and Lake Towuti South Sulawesi. Proceedings of the MLI Annual Scientific Meeting. *The Limnologi Indonesia Community*. Hal. 493 – 507.

Nomosatryo, Sulung., Cynthia Henny, Eti Rohaeti & Irmanida Batubara. (2012). Fraksinasi fosforus pada sedimen di bagian litoral Danau Matano Selatan. Prosiding Seminar Nasional Limnologi VI tahun 2012. *Pusat Penelitian Limnologi LIPI*. Phosphorus fractionation in sediments in the litoral part of Lake Matano Selatan. Proceedings of the 2012 Limnology National Seminar 2012. *Limnology LIPI Research Center*. Bogor. Pp. 493 – 510.

OATES, J. and DAVIES, A., DELSON, i~. 1994. The diversity of living colobines. *Colobine Monkeys: Their Ecology, Behaviour and Evolution*, pp.45-73.

Oktavia, P., and Faoziyah, U. 2016. CHALLENGES IN THE RESTORATION OF LAKE MANINJAU: BRIDGING ACTORS' INTERESTS FOR SUSTAINABILITY. *Proceedings of the 16th World Lake Conference*

Ong, P. & Richardson, M. 2008. *Macaca fascicularis*. *The IUCN Red List of Threatened Species* 2008: e.T12551A3355536. http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T12551A3355536.en. Downlo aded on 31 December 2018.

Payne, A.I., 2015. Asian upland fishes and fisheries. In Freshwater Fisheries Ecology (pp. 377-383). *John Wiley & Sons, Ltd*.

Petr, T. 1995. The present status of and constraints to inland fishery development in Southeast Asia. In: Petr, T. and Morris, M. ed. Indo-Pacific Fishery Commission, FAO Report No. 512 Supplement, FIRI/R512 Suppl., *Rome*, 5–29.

Petr, T. and Morris, M. eds., 1995. Regional Symposium on Sustainable Development of Inland Fisheries Under Environmental Constraints: Bangkok, Thailand, 19-21 October 1994, and Country Reports Presented at the IPFC Working Party of Experts on Inland Fisheries: Bangkok, Thailand, 17-21 October 1994 (No. 512). *Food & Agriculture Org*anisation.

Pfaender, J., Miesen, F. W., Hadiaty, R. K., Herder, F. 2011. Adaptive speciation and sexual dimorphism contribute to diversity in form and functional in the adaptive radiation of Lake Matano's sympatric roundfin sailfin silversides. *Journal of Evolutionary Biology* 24: 2329–2345, http://dx.doi.org/10.1111/j.1420-9101.2011.02357.x

Pfaender, J., Schliewen, U. K., Herder, F. 2010. Phenotypic traits meet patterns of resource use in the radiation of "sharpfin" sailfin silverside fish in Lake Matano. *Evolutionary Ecology*. 24: 957–974, http://dx.doi.org/10.1007/s10682-009-9332-2

Piet, G.J. and Vijverberg, J., 1998. An ecosystem perspective for the management of a tropical reservoir fishery. *International Review of Hydrobiology*, 83, pp.103-112.

Port, J.A., O'Donnell, J.L., Romero- Maraccini, O.C., Leary, P.R., Litvin, S.Y., Nickols, K.J., Yamahara, K.M. and Kelly, R.P. 2016. Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*, 25(2), pp.527-541.

Pratiwi, N. T. M., Rahman, A., Hariyadi, S., Ayu, I. P., and Iswantari, A. 2016. Relationship between states and nutrients load in waters surrounding Samosir Island, Lake Toba, North Sumatra. *Proceedings of the 16th World Lake Conference.*

Putri, D. M., Hadisusanto, S. 2016. Zooplankton diel vertical migration in Lake Laut Tawar, Aceh, Indonesia**.** *Proceedings of the 16th World Lake Conference.*

Qu, C. & Stewart, K.A., 2017. Comparing conservation monitoring approaches: traditional and environmental DNA tools for a critically endangered mammal.

Rahman A, Sentosa A A, Wijaya D. 2012. Sebaran ukuran dan kondisi ikan zebra Amatitlania nigrofascia (Günther, 1867) di Danau Beratan, Bali. *Jurnal Iktiologi Indonesia.* 12 (2):135-145. Distribution of size and condition of the Amatitlania nigrofascia zebra (Günther, 1867) at Lake Beratan, Bali. *Indonesian Iktiologi Journal*. 12 (2): 135-145

Rahman, M. and Prihanto, A.A., 2017. Phosphor-based carrying capacity of Riam Kanan river, South Kalimantan on caged fish farming. *Aquaculture, Aquarium, Conservation & Legislation-International Journal of the Bioflux Society* (AACL Bioflux), 10(5).

Restu, I. W., Kartika, G. R. A., Pratiwi, M. A. 2016. Potential identification of flora and fauna Lake Buyan as basis of tourism development strategy based on aquatic ecosystems. *Proceedings of the 16th World Lake Conference*.

Rivai, A.A., 2018, July. Study of fresh Frigate Tuna (Auxis thazard) quality sold by mobile fish retailer in Makassar. *In IOP Conference Series: Earth and Environmental Science* (Vol. 176, No. 1, p. 012039). IOP Publishing.

Robson, H.L., Noble, T.H., Saunders, R.J., Robson, S.K., Burrows, D.W. and 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(4), pp.922-932.

Sabo, E., Roy, D., Hamilton, P.B., Hehanussa, P.E., McNeely, R. and Haffner, G.D., 2008. The plankton community of Lake Matano: factors regulating plankton composition and relative abundance in an ancient, tropical lake of Indonesia. In *Patterns and Processes of Speciation in Ancient Lakes* (pp. 225-235). Springer, Dordrecht.

Sala, O. E., Vuuren, V. D., Pereira, H., M. *et al*. 2005. Biodiversity across scenarios. In: Ecosystems and Human Well Being: Scenarios, Volume 2, Millennium Ecosystem Assessment (eds Carpenter SR, Pingali PL, Bennett EM, Zurek MB), pp. 375–410. *Island Press, Washington, District of Columbia*.

Santhanam, R., 2015. *Nutritional freshwater life*. CRC Press.

Saragih, B. and Sunito, S., 2001. Lake Toba: Need for an integrated management system. Lakes & Reservoirs: Research & Management, 6(3), pp.247-251.

Sentosa, A.A. and Wijaya, D. 2012. Utilization of natural food by dominant fish in Batur Lake, Bali province. Annual National Seminar IX Fisheries and Marine Research Results, July 14, 2012.

Sentosa, A.A., Wijaya, D. and Tjahjo, D.W.H. 2013. Risk study of the existence of organic lake introduction fish, Bali. *Proceedings of the National Forum for the Recovery and Conservation of Fish Resources - IV*

Snook, Amy. 2009. Investigation of Factors Limiting Pelagic Phytoplankton Abundance and Composition in the Ancient Malili Lakes of Indonesia. Electronic Theses and Dissertations. Paper 368.

Subehi, L., Ismail, S.N., Ridwansyah, I., Hamid, M.A. and Mansor, M., 2018, February. Analysis of the influence of reservoirs utilization to water quality profiles in Indonesia (Saguling–Jatiluhur) and Malaysia (Temengor–Chenderoh) with special references to cascade reservoirs. In IOP Conference Series: Earth and Environmental Science (Vol. 118, No. 1, p. 012025). IOP Publishing.

Sulawesty, F. 2016. Phytoplankton community at littoral zones of Lake Matano in relationship to water quality. Proceedings of the 16th World Lake Conference

Suwelo, I.S., 2004. Spesies Ikan langka dan terancam punah perlu dilindunggi undang-undang. Jurnal Ilmu-Ilmu Perairan Perikanan Indonesia, 12: 153-160. Endangered and endangered fish species need to be protected by law. *Journal of Indonesian Fisheries Aquatic Sciences.*

Suryaningtyas, E. U., Ulinuha, D. Effect of seasonal changes on spatial distribution of bacterial pathogens in Tilapia (Oreochromis niloticus) in Lake Batur. *Proceedings of the 16th World Lake Conference.*

Syandri. H. 1996. Aspek reproduksi ikan Bilih (*Mystacoleucus padangensis* Blkr) dan Kemungkinan Pembenihannya di Danau Singkarak. Disertasi Program Doktor IPB Bogor. Reproductive aspects of

Bilih fish (*Mystacoleucus padangensis*) and Possible Hatcheries in Lake Singkarak. Dissertation of IPB Bogor Doctoral Program.

Syandri. H., Aryani, N., & Azrita. 2013. Distribusi ukuran, reproduksi dan habitat pemijahan ikan bilih (*Mystacoleucus padangensis* Blkr.) di Danau Singkarak. Bawal, 5 (1):1-8. Size, reproduction and spawning habitat distribution of bilih fish (*Mystacoleucus padangensis*) In Lake Singkarak.

Thomas, R., 2005. Fishes and ecological aspects in the southern region of Lake Toba and its associated rivers, Sumatra, Indonesia. *Malayan Nature Journal* 57(1):81-89.

Ushio, M. et al., 2017. Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. *bioRxiv.*

Valdez-Moreno, M., Ivanova, N.V., Elias-Gutierrez, M., Pedersen, S.L., Bessonov, K. and Hebert, P.D., 2019. Using eDNA to biomonitor the fish community in a tropical oligotrophic lake. *PloS one*, 14(4).

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F. and Gaboriaud, C., 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, 25(4), pp.929-942.

von Rintelen, T., von Rintelen, K., Glaubrecht, M., Schubart, C. D., Herder, F. 2012. Aquatic biodiversity hotspots in Wallacea - the species flocks in the ancient lakes of Sulawesi, Indonesia. In: Gower GW, Johnson KG, Richardson JE, Rosen BR, Rüber L, Williams ST (eds), Biotic evolution and environmental change in southeast Asia. *Cambridge University Press,* Cambridge, pp 290–315, http://dx.doi.org/10.1017/CBO9780511735882.014

von Rintelen, T., Wilson, A.B., Meyer, A., Glaubrecht, M. 2004. Escalation and trophic specialization drive adaptive radiation of freshwater gastropods in ancient lakes on Sulawesi, Indonesia. Proceedings of the Royal Society B 271: 2841–2849

Walter, R. P., Haffner, G. D., Heath, D. D. 2009. No barriers to gene flow among sympatric polychromatic 'small' *Telmatherina antoniae* from Lake Matano, Indonesia. *Journal of Fish Biology.* 74: 1804–1815, http://dx.doi.org/10.1111/j.1095-8649.2009.02256.x

Welcomme, R.L. ed., 1988. *International introductions of inland aquatic species* (No. 294). Food & Agriculture Org.

Whitehead, P.J.P. and Nelson, G.J., 1988. Clupeoid fishes of the world: An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shads, anchovies, and wolfherrings. *Food & Agriculture Org.*

Whitten, T., Soeriaatmadja, R.E. and Afiff, S.A., 1996. Ecology of Java & Bali (Vol. 2). Oxford University Press.

Whitten, T. and Damanik, S.J., 2012. *Ecology of Sumatra.* Tuttle Publishing.

Wijopriono, W., Purnomo, K., Kartamihardja, E.S. and Fahmi, Z., 2017. Fishery resources and ecology of Toba Lake. Indonesian Fisheries Research Journal, 16(1), pp.7-14.

Wirjoatmodjo, S., Sulistiono, M.F. Rahardjo, I.S. Suwelo, R.K. Hadiaty. (2003). Ecological distribution of endemic fish species in Lakes Poso and Malili Complex, Sulawesi Island. Founded by the Asean Regional Center for Biodiversity Conservation and the European Commission. Bogor. 30 p.

Wu, Y., Fang, M., Du, L., Wu, H., Liu, Y., Guo, M., Xie, J. and Wei, D., 2014. The nutritional composition and anti-hypertensive activity on spontaneously hypertensive rats of sipuncula Phascolosoma esculenta. *Food & function*, 5(9), pp.2317-2323.

Xenopoulos, M. A., Lodge, D. M., Alcamo, J. *et al*. 2005. Scenarios of freshwater fish extinctions from climate change and water withdrawal. *Global Change Biology*, 11, 1557–1564.

Xie, Y. *et al*., 2016. Environmental DNA metabarcoding reveals primary chemical contaminants in freshwater sediments from different land-use types. *Chemosphere*, 172, pp.201–209.

Yamamoto, S. *et al*., 2016. Environmental DNA as a “snapshot” of fish distribution: A case study of Japanese jack mackerel in Maizuru Bay, Sea of Japan. *PLoS ONE*, 11(3).

Yamamoto, S. *et al*., 2017. Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific Reports*, 7.

Online References

Encyclopaedia Britannica. 2018. Malay Archipelago Islands, Southeast Asia. The Editors of Encyclopaedia Britannica. https://www.britannica.com/place/Malay-Archipelago

FAO. 2018. Chanos chanos (Forsskal, 1775). Cultured Aquatic Species Information Programme. Accessed at: http://www.fao.org/fishery/culturedspecies/Chanos_chanos/en

Fishbase. 2017a. Species in Toba. Accessed at: http://fishbase.org/trophiceco/FishEcoList.php?ve_code=547

Fishbase. 2017b. List of Freshwater Fishes reported from Indonesia. Accessed at: http://fishbase.org/country/CountryChecklist.php?showAll=yes&what=list&trpp=50&c_code=360&sortby=alpha2&ext_CL=on&ext_pic=on&vhabitat=fresh

Four Square. 2018. PT. Gema Rahmi Persada, mine site. https://foursquare.com/v/pt-gema-rahmi-persada-mine-site-kotabangun/4e4a36ac18387418d6eb3668

GPS Coordinates Converter. 2018. Accessed at: https://www.gps-coordinates.net/gps-coordinates-converter

Green Peace. 2016. THE DIRTY WORK OF BANPU: WHAT THAI INVESTMENT IS DOING IN INDONESIA. March 2016. http://www.greenpeace.org/seasia/id/PageFiles/723615/%20The%20Dirty%20Work%20of%20Banpu.pdf

https://www.cabi.org/ISC/datasheet/59751Mining Atlas. 2018. PT Inco, A Nickel Mine in Indonesia. https://mining-atlas.com/operation/PT-Inco-Nickel-Mine.php

https://www.cabi.org/isc/datasheet/82079#3763CD66-9130-4443-BE13-EB052119784D

Insight Guides. 2014. Insight Guides: Explore Bali (Insight Explore Guides) Paperback – 1 Apr 2014

IUCN 2018b. Java Toad. Phrynoidis asper. Accessed at https://www.iucnredlist.org/species/54579/62062983

IUCN. 2018a. Sumatran Surili. Presbytis melalophos. Accessed at https://www.iucnredlist.org/species/18129/7666452

Jordan, L. 2008. Poecilia reticulata (guppy). Invasive Species Compendium. CABI Centre for Agriculture and Bioscience International. Accessed at: https://www.cabi.org/isc/datasheet/68208

LakeNet. 2003a. Lake Profile. Batur. http://www.worldlakes.org/lakedetails.asp?lakeid=10732

LakeNet. 2003b. Lake Profile. Bratan. http://www.worldlakes.org/lakedetails.asp?lakeid=10751

LakeNet. 2003c. Lake Profile. Buyan. http://www.worldlakes.org/lakedetails.asp?lakeid=10737

LakeNet. 2003d. Lake Profile. Matano, Danau, (Matana). http://www.worldlakes.org/lakedetails.asp?lakeid=8694

LakeNet. 2003e. Lake Profile. Melintang. http://www.worldlakes.org/lakedetails.asp?lakeid=10110

LakeNet. 2003f. Lake Profile. Pening, Danau. http://www.worldlakes.org/lakedetails.asp?lakeid=10544

LakeNet. 2003g. Lake Profile. Semayang (Semajang) http://www.worldlakes.org/lakedetails.asp?lakeid=10111

LakeNet. 2003h. Lake Profile. Singkarak, Danau. http://www.worldlakes.org/lakedetails.asp?lakeid=8637

LakeNet. 2003i. Lake Profile. Tamblingan. http://www.worldlakes.org/lakedetails.asp?lakeid=10738

LakeNet. 2003j. Lake Profile. Tawar, Danau. http://www.worldlakes.org/lakedetails.asp?lakeid=10364

LakeNet. 2003k. Lake Profile. Toba (Danau Toba). http://www.worldlakes.org/lakedetails.asp?lakeid=8367

Maddern, M. 2009. Xiphophorus hellerii. Invasive Species Compendium. CABI Centre for Agriculture and Bioscience International. Accessed at: https://www.cabi.org/ISC/datasheet/59751

Mammals of Malaysia. 2018. Feliformia. Accessed at: http://malaysiamammals.myspecies.info/taxonomy/term/1432

Mongobay. 2017. Downstream From a Coal Mine, Villages in Indonesian Borneo Suffer from Water Pollution. http://earthfirstjournal.org/newswire/2017/03/24/downstream-from-a-coal-mine-villages-in-indonesian-borneo-suffer-from-water-pollution/

Siriwardena, S. 2010. Gambusia affinis (western mosquitofish). Invasive Species Compendium. CABI Centre for Agriculture and Bioscience International. Accessed at: https://www.cabi.org/isc/datasheet/82079

Tisnadibrata, I. L., and Wiriyapong, N. 2016. Banpu disputes Greenpeace report on Indonesia operation. Bangkok Post. https://www.bangkokpost.com/print/929361/

U.S. Environmental Protection Agency. 2018. What is Acid Mine Drainage? http://www.sosbluewaters.org/epa-what-is-acid-mine-drainage%5B1%5D.pdf

UN. 2015. "UN Water and Sanitation Best Practices Platform". International Decade for Action "Water for Life" 2005-2015. Web page: http://www.unwaterbestpractices.org/

UNEP. 2018. Technology Needs for Lake Management in Indonesia - Investigation of Rawa Danau and Rawa Pening. http://www.unep.or.jp/ietc/publications/techpublications/techpub-9/index.asp

Versteegh, D. 2010. Danau Matano Nature Recreation Park. Central Sulawesi Danau Matano Nature Recreation Park. Accessed at: https://indonesiatraveling.com/central-sulawesi-danau-matano-nature-recreation-park/

Zoom Earth. 2018. https://zoom.earth/#-0.364351,116.549347,18z,sat

Chapter 6

General Discussion

6.1 Overview of experimental chapters

This thesis explores how eDNA sampling, combined with multi-gene metabarcoding and next-generation sequencing can be used to monitor aquatic biodiversity in tropical regions such as the Malay Archipelago. Firstly, I discuss the development of eDNA sampling and its potential, pros, cons, and future development needed for implementation as a monitoring tool in Southeast Asia with respect to the specific environmental and socio-political issues faced in this region of the world. Secondly, I present a comparison of eDNA capture and storage techniques with a custom protocol for eDNA extraction from an enclosed filter capsule used in combination with a preservation buffer, which our results suggest yield better results than other compared methods, and which can be implemented in eDNA sampling in the tropics. Thirdly, I explore the spatial distribution of eDNA within a small tropical montane lake, and find that taxonomic information as well as OTU richness varies between surface points only 500 m apart and depth points only 2 m apart, suggesting the need for a comprehensive spatial approach to eDNA sampling to detect extant biodiversity in the tropics. Finally, I test the use of eDNA metabarcoding to assess the extant aquatic species of lakes from the Malay Archipelago, and compare taxonomic communities and OTU richness between areas based on a range of habitat variables. I find that altitude, lake area, lake depth and trophic productivity have an effect on community composition as well as OTU richness, and that although eDNA metabarcoding was successful in detecting native, invasive, endemic and rare species - there are many sampling, molecular, and bioinformatic challenges to be overcome before this approach can reliably be used in monitoring aquatic species from biodiversity hotspots such as Southeast Asia.

6.2 Aquatic eDNA collection techniques for biodiversity monitoring

When collecting eDNA samples from tropical lakes, filtering using a broad pore size filter may be an optimal option, as the fine pore filters such as 0.22 µm Sterivex filters easily clog, limiting the volume to be processed. The higher the number of sampling sites, and the higher the volume of water sampled, will improve the likelihood of detecting the extant biodiversity.

Similar to the work completed here in Chapter 2, other recent studies have also compared aquatic eDNA capture and storage techniques. Djurhuus *et al.* (2017) compared polyvinylidene difluoride (PVDF), polyethersulfone (PES), glass fibre (GF), polycarbonate track etch (PCTE) and nanocellulose (NC) filters all of 0.2 µm, and found no significant difference in eDNA results. Another recent study (Majaneva *et al.* 2018) tested four different preservation strategies (on ice, in ethanol, in lysis buffer and dry in silica gel), two filter types

(mixed cellulose ester and polyethersulfone) and found that either dry storage or storage in lysis buffer, and mixed cellulose rather than polyethersulfone gave the most consistent community composition using metabarcoding. Serial filtration for size fractionation could also be beneficial to separate different types of eDNA, capture different sections of a biological community, or remove larger organic particles (Alawi *et al*. 2014; Bass *et al*. 2015), particularly relevant to the highly turbid waters of Southeast Asia, where agricultural runoff is prevalent. Therefore, the ecological question, environmental habitat, specific environmental sample type, and specific target organisms must all be considered when planning the method of eDNA isolation.

6.3 The use of eDNA in wildlife and biodiversity monitoring in the Malay Archipelago

The Malay Archipelago of Southeast Asia has unique challenges relating to invasive species, river impoundment, overexploitation and pollution (discussed in the draft manuscript in Appendix 3). The use of aquatic eDNA monitoring could provide valuable information regarding the presence and distribution of particular species, and predict their response to such threats.

Flow modification through river impoundment by e.g. hydrological dams for hydropower present a huge problem for freshwater species in Southeast Asia, where 98 dams are planned for construction by 2030 in the Mekong basin alone, with an additional 371 dams already operational or under construction. An increase of this magnitude would require a 19-63% expansion of agricultural land to preserve regional food security in the face of projected fishery loss (Winemiller *et al*. 2016). Hydropower dams alter natural flow regimes with consequences for water temperature, nutrient loads and sediment transport downstream, and contribute to terrestrial and aquatic species and habitat loss, reduction of fishery yield and deter fish migration (Stone 2011; Winemiller *et al*. 2016, Welcomme *et al*. 2016). Pfleger *et al*. (2016) investigated the impact of dams and barriers on the critically endangered Alabama Sturgeon (*Scaphirhynchus suttkusi*) and near threatened Gulf Sturgeon (*Acipenser oxyrinchus desotoi*) using eDNA, and found both species remained upstream of passage barriers. Consequently, the authors recommended that the removal of the barriers to passage would aid in the conservation of these species. One of the high priority topics in which aquatic eDNA metabarcoding in Southeast Asia could be implemented therefore, is in investigating the impact of hydrological dams on biodiversity.

Water pollution is also a major threat to Southeast Asian freshwater habitats. Overloading of nutrients from agricultural fertilizers can cause harmful algal blooms, which can increase cyanotoxins and cause harmful bioaccumulation in aquaculture fish such as tilapia (Greer *et al*. 2017). Healthy freshwater ecosystems act as natural pollutant filters (Chowdhury *et al*. 2016, Cochard 2017), which can be more economically effective than industrial water filtration plants (Collen *et al*. 2014). Another important avenue for aquatic eDNA metabarcoding studies in Southeast Asia therefore, is in understanding the impact of pollutants on freshwater biodiversity, and how more biodiverse habitats can act as natural pollutant filters.

For this type of research to be conducted in Southeast Asia, high level molecular infrastructure such as fully equipped PCR free laboratories and sequencing centres are necessary. This either requires the presence of such facilities within the Southeast Asian country, or the export of raw samples or DNA extracts to laboratories in other countries. Exporting samples from some Southeast Asian countries can be challenging. The Indonesian research permit process is strict, extensive and complicated, and the export of samples extremely difficult. If infrastructure does not exist, and international collaborations are necessary to implement eDNA metabarcoding, it is important for non-Indonesian researchers to first establish thorough connections with Indonesian governmental bodies to navigate the appropriate permit steps.

6.4 Future perspectives on eDNA metabarcoding

There is a much-repeated need for sequencing of mitochondrial barcode regions, or preferably whole mitochondrial genomes, to populate genetic databases and consequently improve the accuracy of species assignment, allowing species detection from eDNA using popular barcoding primers (Ishige *et al*. (2017). Sequencing larger barcodes or whole mitogenomes could be beneficial to eDNA metabarcoding studies for several reasons. Larger barcodes, or whole mitogenomes could allow haplotype counts to be used for better abundance estimates (Stat *et al*. 2017) rather than the approach of analysing read counts which is prone to bias. In addition, databases for genes other than COI are severely lacking in taxonomic coverage, and so there is an urgent need for reference libraries of other standard barcodes to be expanded to increase the ability of short universal markers to match with information in genetic databases (Leray and Knowlton 2015, Creer *et al.* 2016, Coisacc *et al.* 2016). It has also been suggested that investigators involved in metabarcoding studies should plan to barcode representatives of their local biota as a part of their projects (Porter and

Hajibabaei, 2018). Coissac *et al.* (2016) argue that there is a need to continue, and in fact accelerate global efforts to build not only the DNA barcode reference library of standard barcodes, but also that of what the authors refer to as 'extended barcodes' to strengthen the standard barcoding approach. Extended barcodes (see glossary), which can be created using genome skimming, provide higher phylogenetic signal than standard barcodes, providing a bridge between standard and metabarcoding studies that usually use particular target regions, and provide a way to circumvent the often-biased approach of target enrichment with PCR (Coissac *et al.* 2016).

In a recent study comparing shotgun sequencing and metabarcoding using mock communities of freshwater macroinvertebrates, metabarcoding was less consistent than shotgun sequencing, and failed to recover some species with higher abundances, whilst shotgun sequencing results provided highly significant correlations between read number and biomass in all but one species (Bista *et al*. 2018). However, whole genome shotgun sequencing in metagenomic studies requires a large proportion of data to be discarded, and a huge increase in sequencing output is required, resulting in a decrease in sample throughput compared to metabarcoding. For example, Stat *et al*. (2017) performed shotgun sequencing on marine eDNA samples and of the 22,300,000 sequencing reads obtained, only 14% (3,122,000) could be assigned to anything using Blastn, and only 2.4% of those reads (74,928) had assignments matched to eukaryotes, with 94.5% assigned to bacteria, and 3% to viruses. Furthermore, of the 2.4% of eukaryotic reads, only 1.2% of these reads (899) were assigned to fish, meaning that of the original 22.3 million reads, only 0.004% were assigned to fish, demonstrating the unsuitability of shotgun sequencing in monitoring aquatic vertebrates. This can be alleviated by combining shotgun sequencing with DNA capture array technology (Liu *et al.* 2016) to target specific organelles, and to hybridize and extract specific genomic regions, subsequently reducing the size of the genomic target and increasing the number of samples (Creer *et al.* 2016). The MinION™ sequencer continues to show promise as an option in 'benchtop' genomics, and recently proven its potential in metagenomics. Brown *et al*. (2017) used the MinION™ to sequence three types of low-complexity synthetic communities from four bacterial species, a community with one relatively rare (1%) and three abundant (33% each) components, and a mixture of genomic DNA from 20 bacterial strains. They generated accurate taxonomic assignment of high-quality reads from the MinION approaching 99.5% and inferred community structure mostly mirrored the known proportions of these synthetic mixtures.

Clarity on the effect of bioinformatics processing of samples will provide a baseline of information for eDNA studies to make standardised decisions regarding e.g. filtering for sequences found in certain numbers of replicates, or a certain copy number which can drastically change the final outcome of species lists (Evans *et al*. 2017a; Leray and Knowlton, 2017). This type of filtering e.g. removing reads found with < 10 copies can account for errors created by random sampling of rare sequences during the Illumina sequencing process (Leray and Knowlton, 2017).

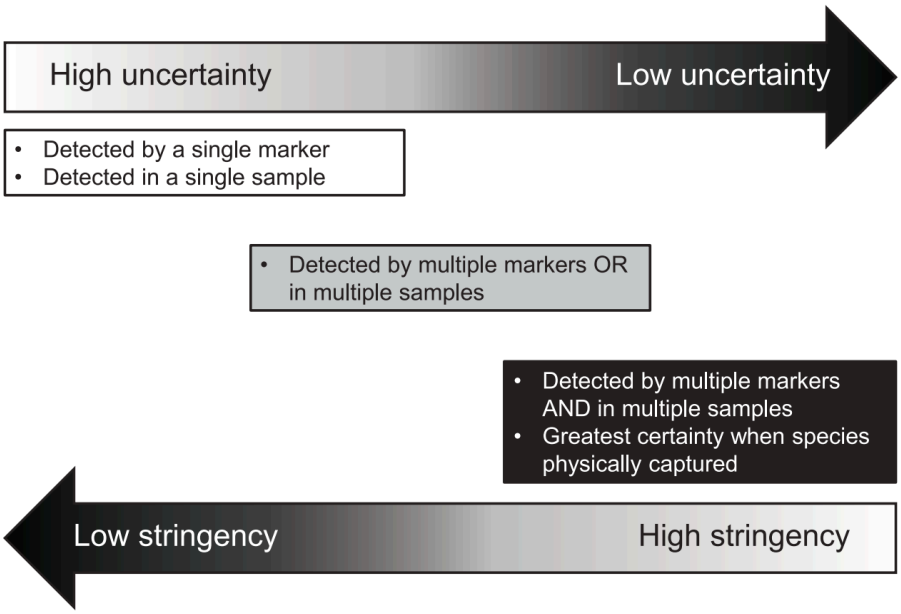


Figure 6.1 Bioinformatic stringency and taxonomic certainty. Conceptual diagram illustrating the relationship between bioinformatic stringency and strength of certainty about the presence of species detected with eDNA metabarcoding (Evans *et al*. 2017).

Evans *et al*. (2017) demonstrated this pattern, by metabarcoding aquatic eDNA targeting fish communities with 'low', 'moderate' and 'high' bioinformatic stringency producing 21, 15, and 8 fish OTUs (Operational Taxonomic Units, used as approximations for species, see Glossary page 17) compared to the 10 from direct observations. Alberdi *et al*. (2017) go further and test over 2,000 combinations of molecular and bioinformatic replication and filtering, and showed that OTU number was greatly affected by the number of PCR replicates, how samples are filtered across them, sequence copy number, and OTU clustering threshold. However, Lahoz-Monfort *et al*. (2016) suggest that removal of single PCR detections as an ad-hoc filtering approach to account for false negatives or positives results in biased estimation of occupancy, detectability and false positive rates, and that instead, prior

information or additional data collection using other methods to perform 'site occupancy-detection modelling' should be incorporated. Giguet-Covex *et al*. (2014), for example, suggest that a sequence should only be considered if confirmed by at least two independent PCRs, whilst those detected in only one replicate should be discarded or considered dubious, although this can drastically reduce the number of species recorded (Evans *et al*. 2017a). The interpretation and implementation of this information is yet to be consistently applied in eDNA research.

When considering options for OTU clustering approaches, the appropriate choice is complex due to either a lack of marker variation between individuals or species, and the creation of spurious OTUs due to read errors from sequencing artefacts and chimeras (Sokal, 1963; Sneath and Sokal, 1973). These issues can artificially inflate biodiversity estimates, invalidate rarefaction curves for alpha and beta diversity estimators, and disrupt the topology of phylogenetic trees. In fact, contrary to the recognized 97% norm, a recent study suggests a 99-100% threshold is more appropriate (based on analysis of microbial communities, with 99% found to be best for full length 16S sequences and 100% found to be best for the V4 hypervariable region) (Edgar, 2018). However, the approach used in this study was criticised, due to the consensus that the 16S gene is not suitable for delineating bacterial species, and, using a higher threshold risks splitting sequences from the same genome into different OTUs (Schloss, 2018). A recently suggested solution is the use of exact amplicon sequence variants (ASVs) to replace OTU clustering altogether in marker gene data analysis (Callahan *et al*. 2017). When using OTU clustering approaches, benchmarked algorithms for quality control, de-noising, chimera removal, OTU picking, subsampling, appropriate distance levels to define OTUs, and a robust method for taxonomic assignment with statistical inference are all required (Cristescu, 2014; Leray and Knowlton, 2017; Wilkinson *et al*. 2018). Initial OTU delineation creates an estimation of species diversity, providing a framework for subsequent taxonomic amendments, but the use of exact ASVs may be a better option for more accurately assigning species.

6.5 Additional work

There was additional work undertaken during the investigations of this PhD which is not featured in this thesis.

In 2014, the first field expedition was in Peninsular Malaysia, where a mesocosm experiment was set up to test eDNA degradation within the environmental habitat variables

of Malaysia, and several lake sites were sampled using an ethanol precipitation method. The mesocosm experiment used 12 x 45 litre buckets dug into the ground, with an equal combination of 0, 1, 2, and 4 fish present. Water samples were collected before and after fish were added, and after fish were removed at multiple time points using 15 ml centrifuge tubes. This experiment was intended to test the accumulation of eDNA over time, and the degradation of eDNA after the biological source was removed. However, amplification was observed from some samples taken from 0-fish buckets, and no amplification was observed from buckets containing fish. It appeared therefore that the results were unreliable, and so this experiment was omitted from the thesis. These results could have occurred for several reasons. Although lids were placed on the buckets each night and removed each morning to prevent the nightly monsoon rainfall from causing flooding, there were occasions when it was not possible to place lids on the buckets during the day when a sudden rainfall occurred. This may have caused the resident fish to escape or move between buckets. Also, Kingfisher birds were observed close to the buckets, and may have hunted the experimental fish, removing them from experimental buckets. It is also possible that laboratory induced errors caused amplification or non-amplification of unexpected samples. If this experiment were to be repeated, a net covering could have been used over the mouth of the buckets to prevent external predator influence, and a waterproof cover installed above the experimental area to prevent the interaction of heavy rainfall. Additionally, rather than one 15 ml sample collected per replicate bucket at each time point, multiple samples would have allowed for comparison of results and allow for laboratory induced error.

The sampling of lakes in Peninsular Malaysia using ethanol precipitation was unsuccessful at the point of amplification. The gel electrophoresis images of PCR products amplified from these eDNA samples were either very faint or absent, and Qubit results were very low. Sequencing of these samples showed little to no amplification of freshwater biodiversity, and so this part of the study was omitted from the thesis. This could have been the result of a lack of sample replicates per lake, or a lack of water volume, as one x 15 ml sample was collected at roughly 1 km points around the edges of each lake. If this study were to be repeated, at least three replicate samples would be collected per sampling point, and based on the results from Chapter 4, the number of sampling points increased to every 200 m around the edge of the lake. It may also be possible that samples were degraded during the shipment from Malaysia to Bangor University, and then to Copenhagen University where the majority of the laboratory work was processed. Samples might have degraded at several stages of the sampling process 1) after filtering before being stored in the freezer, 2) after being stored in

the freezer before extraction 3) after extraction when being transported from Indonesia to Bangor and then again to Denmark. This may have been prevented if the eDNA samples (which were filtered onsite) were immediately stored in a preservation buffer, as we suggest in Chapter 2. This is in contrast to Valdez-Moreno *et al*. (2018) who extracted all samples within 48 hours of filtering, and had good results, detecting 75 species of vertebrates including 47 fishes, 15 birds, 7 mammals, 5 reptiles, and 1 amphibian.

Samples were repeatedly frozen and defrosted to test and develop the methods used to amplify them, which may have compromised their DNA yield. If this study were to be repeated, eDNA extracts would be diluted into several sub-extracts stored in PCR strips, immediately frozen, and additional 'test' samples collected to test and develop molecular approaches such as PCR conditions.

6.6 Limitations and suggestions for future improvements

Aquatic environmental DNA metabarcoding - much like other techniques used to survey biodiversity - is an imperfect solution to a naturally complex challenge. There are biases at every level of the pipeline, introduced through capture technique, primer choice, PCR stochasticity, sequencing ability and OTU clustering. However, metabarcoding does generate a vast amount of taxonomic information from relatively few samples, and when molecular pipelines are thoroughly tested and developed, provides a fast and reliable method of monitoring more biodiversity than could be detected using traditional methods (e.g. Thomsen *et al*. 2012; Dejean *et al*. 2012; Mächler *et al*. 2014; more examples in the General Introduction section 1.5).

For eDNA capture, this study would have been improved by the use of a preservation buffer injected into the Sterivex capsules. At the time of sampling, the initial Qubit results from Chapter 2 (Spens *et al*. 2017) suggested that freezing the filters generated better DNA yields, and so this option was chosen for the main sampling trip in 2015 which occurred before the qPCR results were generated. The method of freezing samples was chosen to be logistically simpler in the field, so that injecting a buffer using pipettes was not needed. However, based on Spens *et al*. (2017), and experience in the field with regards to the logistics of accessing suitable freezers and keeping samples cold enough during transport in tropical climates, the use of a buffer injected into the filter, such as Longmire's solution, is strongly advised. Other tropical aquatic eDNA studies have had success when using Sterivex filters combined with, for example, RNALater as a storage buffer (Ishinge *et al*. 2017).

In this study, 0.22 µm polyethersulphone filters were used (Sterivex-GP Pressure Filter Unit SVGPL10RC), which may not have been the optimal filter material for sampling such water such as that of Lake Tamblingan which was fairly eutrophic and is therefore likely to contain myriad microorganisms. As the 0.22 µm filter clogged easily when filtering some of the more turbid lakes such as Semayang and Melintang, a broader pore size filter would have allowed a greater volume to be sampled. A recent study has suggested that 0.8 µm filters may be the optimal size (Li *et al*. 2018), as there was little difference between 0.45 µm and 0.8 µm in DNA yield and probability of species detection, but using a 0.8 µm filter reduced filtration time by 36%. In addition, 0.8 µm and 1.2 µm filters actually performed better in terms of correlation between read counts and fish abundance. This could then allow more water to be filtered per filter, increasing the probability of species detection by allowing a greater volume of eDNA to be concentrated, and perhaps more areas of the same lake to be sampled. If eDNA is mainly composed of whole cells, then these larger pore size filters should be sufficient to capture the genetic information contained within aquatic eDNA samples. This study also performed a pre-filter step using 20 µm filters, then filtering the expelled water through a 0.45 µm, and found that this reduced filtration time by around 50%. Another option could have been to be flexible with the amount of water filtered, e.g. anything between 500 ml and 1.5 L depending on how quickly filters become clogged, as has been employed by Agersnap *et al*. (2017). This may have allowed more water to be filtered from very clear lakes such as Lake Matano, and possibly increased the detection probability of the extant aquatic species.

For the molecular workflows, the addition of a human blocker oligonucleotide primer may have aided in preventing human amplification. Although the use of human blockers in combination with eDNA metabarcoding is not always employed, this approach may have reduced the human contamination observed. A human blocker to complement the molecular workflow of the 12S Teleost primers (Valentini *et al*. 2016) has already been designed - (teleo_blk: ACCCTCCTCAAGTATACTTCAAAGGAC-SPC3I) (Valentini *et al*. 2016), and used by Sigsgaard *et al*. (2017). However, when considering the addition of a human blocker to the PCR recipe, it is important to note that such an approach may not eradicate the presence of human DNA, and may decrease the detectable diversity (Piñol *et al*. 2015). For example, Thomsen *et al.* (2016) observed human DNA in all samples, although a human blocker was used in the PCR set up, designed to complement the 12S Teleost primers (also used herin) (Valentini *et al*. 2016).

Another challenge in the molecular workflow of aquatic eDNA analysis, is that when working with such low quantity DNA as eDNA it seems somewhat inevitable that a certain level of contamination may be expected, even when taking significant precautions to limit such extraneous DNA. Thomsen *et al*. (2016) observed human, chicken, rock pigeon, duck, and lionfish DNA in their Greenlandic marine eDNA samples, which are all likely to be false positive results. Human DNA, DNA from positive controls, and DNA from the Saola (*Pseudoryx ngethinhensis*) was observed in the OTU tables from Chapter 4 and 5. The Saola DNA was only found from samples and negative controls extracted at the laboratories at the Natural History Museum of Denmark, and so it is likely that this was from laboratory based contamination carried over by genome sequencing being performed on this species by another researcher. The reporting of such contamination should be standard practise, and if carefully considered, should not impact the interpretation of genuine eDNA data, as was done by e.g. Thomsen *et al*. (2016) and Stat *et al* (2017). Based on the results of this thesis, although aquatic eDNA is considered 'modern' DNA, it is of such low quantity, and easily prone to contamination, that it would be beneficial to work in near-ancient DNA laboratory conditions, and perhaps the stringency of laboratory rules should be based on the quantity of DNA in the sample, rather than the age. For example, iDNA from leeches is carried within the prey blood inside the leech at high quantities, and so should not need to be processed in strictly clean labs, but filters containing trace DNA from water samples are no more likely to cause contamination to clean lab environments than, for example, fragments of ancient bone dug from the ground, or ancient lake sediment cores. A high degree of human DNA was amplified from most of the lake samples from this thesis, which may have been prevented if samples were processed in a lab area where no human DNA samples (or any samples at all) were being amplified, and could possibly contribute PCR product contamination.

Another molecular limitation may have been in the pooling of aquatic eDNA samples from the same lake, which may result in the loss of eDNA found in low copy numbers, subsequently yielding a lower species richness (Sato *et al*. 2017). Therefore, it may be a better approach to individually extract different samples from within the same lake, which would also allow the estimations of the means and standard deviation amongst the replicates. However, Sato *et al*. (2017) who explore this suggestion refer to reads which contribute < 0.05% of each sample which could be spurious assignments.

This study could have been improved further by keeping eDNA extracts of ecological replicates separate, and PCR replicates separate, rather than combining eDNA extracts of ecological replicates before PCR (although this would triple the molecular work load and cost

of sequencing). On the other hand, other eDNA studies have combined PCR replicates, as opposed to combining ecological replicates, although the disadvantage of this approach is the lack of ability to remove spurious sequences only found in one PCR as was used in this approach for Chapter 4 and 5. However, based on these results, if a sequence has entered the sample through lab contamination, it is likely to be of such high read count that the source is obvious, and these OTUs can be removed (such as the human and Saola DNA observed herein).

Due to the high number of samples being processed for metabarcoding, the eDNA extracts were set up for PCR in chronological order of collection per lake. However, this is not a randomized and/or balanced approach. It has recently been suggested that molecular methods should report the detailed design of sample processing in the laboratory, as this may strongly influence the interpretability of results where confounding effects may occur (Bálint *et al*. 2018).

At the bioinformatic step of the workflow, another limitation of this study was the interpretation of reads assigned to species which were used as positive controls. Several tissue samples were extracted from Malaysian and Indonesian fish to validate that primers were able to amplify these targets (see Chapter 3: Universal Methods). To use up the remaining PCR wells, these extracts were added individually. A better option would have been to create a mock community using these extracts, by diluting them down to roughly that of the eDNA samples, and combining at equimolar ratios. This would have prevented such high read abundances sequenced in the positive control samples, and prevented the overspill contamination observed in other samples. It is impossible to know, therefore, whether the reads observed in the lake eDNA samples were genuine, or overspill from the positive controls, and so all OTUs assigned to these species were removed from the analysis. Based on this problem, I advise future studies to not sequence target species' DNA on the same sequencing run as eDNA samples which may contain these target species, and to dilute positive control samples down to roughly that of eDNA samples, and to store positive control DNA extracts in a separate box to eDNA samples.

As a short fragment of 12S (60-80 bp) and 16S (90 bp) was analysed, some amplicons gave 100% full-length matches to multiple species, as is to be expected. An *ad-hoc* species-level identification could be made in some cases where only few species exist within a genus, based on the known geographic range of those taxa. However, future metabarcoding work would benefit from markers of longer fragment lengths which may allow the delineation of fish species with low variability in the 12S or 16S marker region used.

It is likely that there are false negatives from the eDNA samples described in Chapter 4 and 5, as there were many more species described from the literature than were detected from the samples. It may be that the bioinformatic filtering approach employed was too stringent. For example, when considering OTU tables based on filtering a minimum of 2 copies of DNA per PCR replicate, rather than 20 (as was used for the 12S and 16S data), there are species which are likely to be real, local, eDNA signals which were then removed by the filtering process. In the 16S data, the Convict Cichlid (*Archocentrus nigrofasciatus*) appears in sample 'DTAMS9' from Lake Tamblingan in Bali, which is supported by the literature (Candrawan, 2015) but this species was only detected from this sample, and only appears with 9 copies found in 2/3 PCR replicates (with 100% identity and query cover BLAST result to this species assignment), and so was removed. Additionally, if 1/3 PCR minimum threshold was used, many more species would probably be detected, although this could compromise the results in terms of false positives.

There were other issues related to BLAST and NCBI database problems. Some hits show low identity, but can have consistent results to a particular taxonomic assignment, giving more confidence in that assignment. For example, for 12S data, the sequence cccctgtcaaacgcacaaaaatatataataaactagcactcgacaagaggaggcaagtcgtaa (OTU 102) with 767 reads before normalisation in a sample from the Malaysian reservoir lake, Tasik Chendorah, returned 'no hit' from the MEGAN assignment when using megablast, but when run using blastn showed a list of matches. Of these matches, the top four hits were to fish in the *Crossocheilus* genus, with perfect query cover but relatively low identities (91-92%), possibly due to a lack of sequences in the database. Additional hits matched to other fish from Asia within the Cyprinidae family, (in decreasing order of E value) *Epalzeorhynchos frenatus*, (Rainbow sharkminnow), *Lobocheilos melanotaenia* (cyprinid fish from the Mekong), *Rectoris posehensis* (cyprinid fish from Asia), *Thynnichthys thynnoides* (Tiny scale barb), *Epalzeorhynchos bicolor* (Red-Tailed Black Shark) and *Ptychidio jordani* (the ratmouth barbel).

In contrast, other 'no hit' sequences also included what appear to be chimeras, in which the top sequences are highly inconsistent and of low quality. For example, an OTU from the 12S data (OTU 100), with 936 reads in a sample from the Indonesian lake, Danau Singkarak, with the sequence cccccgccccactttaaatataaagccttaaataaatctaaacacacccgcaaggggaggcaagtcgtaa returned 'no hit' from a megablast MEGAN assignment, but returned a list of matches when using blastn.

These matched (in decreasing order of E value) to *Silhouettea* (genus of marine goby from the Gobiidae family), *Bacillus glycinifermentans* (bacteria species), *Favonigobius gymnauchen* (Sharp-nosed sand goby from the Gobiidae family), *Mastacembelus mastacembelus*, (Euphrates spiny eel, from the spiny eels family - Mastacembelidae), *Microphis brachyurus*, (Short-tailed pipefish from the seahorse and pipefish family - Syngnathidae), *Anoxypristis cuspidata* (knifetooth sawfish, from the sawfish family - Pristidae), *Gobiodon histrio* (marine goby species - road-barred goby, from the Gobiidae family), *A. cuspidata* (as above), Fiji disease virus (a plant virus) and *Salamandra atra* (the alpine salamander).

Some hits match to a particular species, when it may be the case that there is not enough variation in the gene for resolution between very closely related species, e.g. MEGAN assignment initially showed *Pristolepis grooti* (Indonesian Leaffish) in samples in Lake Melintang and Lake Chenderoh, however it is its close relative *P. fasciata*, that is listed in the literature at Lake Melintang, which also came up in the BLAST search, although neither species matched 100%. These are the only two species in the *Pristolepis* genus in Indonesia and Malaysia (according to Fishbase). There were two OTUs created matching *P. grooti*, one which was only found in Lake Melintang and one which was only found in Lake Chenderoh, implying that these may be separate species, or at least genetically distinct forms of the same species.

In other cases, there were 100% match to many different species, such as OTU20 in the 12S dataset, which matched to 100% to *Sarotherodon galilaeus* (Mango Tilapia), *Oreochromis niloticus*, (Nile Tilapia), *Oreochromis aureus* (Blue Tilapia), and *Sarotherodon melanotheron* (Blackchin Tilapia), as these species are closely related, and must have little genetic variation within the 12S gene. Additionally, some species listed in publications had taxonomic ambiguity, for example, when searching the literature for '*Cyclocheilichthys de Zwani*' recorded in Mardiah *et al*. (2016), Google Scholar returned no matches, although there were 438 Google results, mostly in Bahasa Indonesian. A search on Fish Base for this species returned n = 1 of a 'Possible Scientific Name' of *Cyclocheilichthys dezwaani* (Weber & de Beaufort, 1912), for which the current true classification is *Cyclocheilichthys armatus* (Kottelat and Lim, 1996). A similar problem occurred with another species listed in this paper as '*Puntius shwanefeldi*' but is actually now classified as '*Barbonymus schwanenfeldii*'. Futhermore, there are what only appear to be spelling mistakes in some of the literature, such as in De Silva (1987), who records '*Ophicephalus seriatus'* in 'Table 2. The major reservoirs of Indonesia and their characteristics.' although in a Google search,

there are only two results for this species, and it is highly likely the author is referring to '*Ophicephalus striatus'*, the now outdated name for *Channa striata*.

Some sequences had very poor BLAST results, but were still assigned to a species or genus by MEGAN. For example, in the COI data, OTU 476 was assigned to '*Leptodactylus* sp' (a genus of Neotropical leptodactylid frogs), but a query cover and identity score of only 51% and 80% were observed, and the subsequent hits matched to a mushroom, another frog species, a weevil species and a bird species. In these instances, a new assignment was given of 'insufficient hit', and were removed from the analysis.

Similarly, poor BLAST results were also observed where there were many hits for the same species, such as from OTU 376, which was assigned to 'Rhacophorinae' but with only 19% query cover and 90% identity, but with highly consistent matches to *Kurixalus bisacculus* (a Southeast Asian frog species).

There are methods, such as the recently created LULU algorithm (Frøslev *et al*. 2017) which remove erroneous OTUs by combining information on sequence similarity and co-occurrence patterns, without discarding rare but real OTUs. This may have allowed an improvement in the species assignments of the OTUs generated.

When interpreting the results of the OTU tables generated by the DAMe pipeline after MiSeq sequencing, there were challenges faced in understanding accuracy of the species assignments generated by MEGAN. The consistency of species names of local fish species proved problematic. Either old references refer to a name now not used (e.g. Green *et al*. 1978 refers to *Sarotherodon mossambica*, now *Oreochromis mossambicus*), or slight differences were observed in names such as *Pristolepis fasciatus* (synonym) and *Pristolepis fasciata* (accepted name) (Fishbase, 2018). Old species names appear in BLAST searches which are now out of use. Furthermore, some Indonesian publications only refer to Indonesian, rather than Latin names of species, and when translating publications in Bahasa Indonesia (the Indonesian language) to English, the names may have a different meaning in each language. For example, 'Ikan Zebra' (used in Candrawan, 2017) in Bahasa Indonesia, literally translates to 'Fish Zebra', or 'Zebra Fish', which in English would refer to Zebrafish (*Danio rerio*), but in Bahasa Indonesia refers to the Zebra Cichlid, otherwise known in English as the Convict Cichlid (*Amatitlania nigrofasciata*). Other studies only refer to the common name such as 'Ikan Lele' or 'Lele Dumbo' which only refers to genus level (*Clarias*), not specific species, e.g. Negara *et al*. (2015).

6.7 Implications of this work for eDNA monitoring and future suggestions

The lack of species level assignments possible from OTUs generated in Chapter 4 and 5, (particularly of cyprinid fish species), highlights the desperate need for an increase in barcoding work, and possibly description of new species, in the mega-diverse region of the Malay Archipelago. Barcoding work in Europe and North America is disproportionately conducted compared to Southeast Asia, where biodiversity is significantly higher, and anthropogenic threats such as deforestation, river impoundment and pollution are more immediate. If future molecular work focused on barcoding of a range of mitochondrial markers, and ideally whole mitogenomes, the relevance of metabarcoding work such as that presented in this thesis would be significantly improved.

In Chapter 5, eDNA samples were collected along a transect by subsampling every 500 m for 2.5 km. It would be interesting for future work to compare the taxonomic information generated from many subsamples pooled into one large sample, with processing each subsample separately. It is likely that processing each sample separately would increase the taxonomic and ecological information generated, but as with most ecological surveys, this would be limited by time and resources.

The apparent highly localised nature of eDNA, further illuminated by the results of the data presented in this thesis, has implications for biodiversity monitoring in Southeast Asia. The monitoring of waters above and below hydroelectric dams, for example, could provide useful information in SEA where hydropower dams present an increasing problem for fish populations by blocking migration. There are 98 dams planned for construction by 2030 in the Mekong basin alone, with an additional 371 dams already operational or under construction, with catastrophic results predicted for aquatic biodiversity (Winemiller *et al.* 2016).

6.8 Concluding remarks

Overall, this work has attempted to provide evidence of the applicability of aquatic eDNA metabarcoding for monitoring biodiversity in tropical environments for the improvement of conservation biology, monitoring of invasive species, and ecosystem level analysis. I have demonstrated here that the detection of biodiversity from tropical lakes using aquatic eDNA metabarcoding is possible, and present a sampling and molecular method to do so. I show that eDNA is heterogeneously distributed with a tropical lake, and suggest that sampling approaches include a fine scale approach when aiming to assess tropical diversity. I provide evidence of the applicability of aquatic eDNA metabarcoding in the tropics, by recovering native, invasive and rare species of conservation concern from relatively few

samples. Finally, I show that this approach can uncover ecosystem wide patterns driving species communities, based on a range of habitat variables. Aquatic eDNA for biodiversity monitoring will be improved with further barcoding work, especially whole mitochondrial genomes to populate genetic databases to monitor this mega-diverse region of the world.

6.11 References

Agersnap, S., Larsen, W.B., Knudsen, S.W., Strand, D., Thomsen, P.F., Hesselsøe, M., Mortensen, P.B., Vrålstad, T. and Møller, P.R., 2017. Monitoring of noble, signal and narrow-clawed crayfish using environmental DNA from freshwater samples. *PloS one*, 12(6), p.e0179261.

Alawi *et al.* 2014. A procedure for separate recovery of extra-and intracellular DNA from a single marine sediment sample. *J. Microbiology*. Method, 104, pp. 36-42

Alberdi, A., 2016. Scrutinizing key steps for reliable metabarcoding of environmental samples.

Bálint, M., Márton, O., Schatz, M., Düring, R.A. and Grossart, H.P., 2018. Proper experimental design requires randomization/balancing of molecular ecology experiments. *Ecology and evolution*, 8(3), pp.1786-1793.

Bass, D. *et al.*, 2015. Diverse Applications of Environmental DNA Methods in Parasitology. *Trends in Parasitology*, 31(10).

Breitwieser, F.P. and Salzberg, S.L., 2016. Pavian: Interactive analysis of metagenomics data for microbiomics and pathogen identification. *bioRxiv*, p.084715.

Brown, B.L., Watson, M., Minot, S.S., Rivera, M.C. and Franklin, R.B., 2017. MinION™ nanopore sequencing of environmental metagenomes: a synthetic approach. *GigaScience*, 6(3), pp.1-10.

Callahan, B.J. *et al*. 2016 DADA2: high-resolution sample inference from illumina amplicon data. *Nature Methods*. 13, 581–583.

Candrawan, I.B.G., 2017. Kosmologis Masyarakat Hindu Di Kawasan Tri Danu Dalam Pelestarian Lingkungan Hidup. Jurnal Ilmu Agama Dan Kebudayaan, 14(27).

Candrawan, I.B.G., 2017. Cosmological Hindu Society In Tri Danu Area In Environmental Conservation. Journal of Religion and Culture, 14 (27).

Carøe, C. *et al*., 2017. Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution*.

Chowdhury GW, Zieritz A and Aldridge DC. 2016. Ecosystem engineering by mussels supports biodiversity and water clarity in a heavily polluted lake in Dhaka, Bangladesh. *Freshwater Science* 35: 188–199.

Clarke, L.J. *et al*., 2014. Environmental metabarcodes for insects: In silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, 14(6).

Cochard R. 2017. *Coastal Water Pollution and Its Potential Mitigation by Vegetated Wetlands: An Overview of Issues in Southeast Asia.* Redefining Diversity and Dynamics of Natural Resource Management in Asia, Volume 1. Elsevier Inc.

Corlett, R.T., 2017. A Bigger Toolbox: Biotechnology in Biodiversity Conservation. *Trends in Biotechnology*, 35(1).

Cowart, D.A. *et al*., 2015. Metabarcoding is powerful yet still blind: A comparative analysis of morphological and molecular surveys of seagrass communities. *PLoS ONE*, 10(2).

Deagle, B.E. *et al*., 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology letters*, 10(9).

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E. and 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*(4), pp.953-959.

Davy, C.M., Kidd, A.G. & Wilson, C.C., 2015. Development and validation of environmental DNA (eDNA) markers for detection of freshwater turtles. *PLoS ONE*, 10(7).

Djurhuus, A., Port, J., Closek, C.J., Yamahara, K.M., Romero-Maraccini, O., Walz, K.R., Goldsmith, D.B., Michisaki, R., Breitbart, M., Boehm, A.B. and Chavez, F.P., 2017. Evaluation of

Filtration and DNA Extraction Methods for Environmental DNA Biodiversity Assessments across Multiple Trophic Levels. *Frontiers in Marine Science*, 4, p.314.

Doi, H. *et al*., 2015. Droplet digital polymerase chain reaction (PCR) outperforms real-time PCR in the detection of environmental DNA from an invasive fish species. *Environmental science & technology*, 49(9).

Doi, H. *et al*., 2015. Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. *PLoS ONE*.

Drummond, A.J. *et al*., 2015. Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience*, 4(1).

Edgar, R.C., 2018. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics*, 1, p.5.

Evans, N.T. *et al*., 2017. Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, p.cjfas-2016-0306.

Everett, M. V., Park, L. K. 2017. Exploring deep-water coral communities using environmental DNA. *Deep Sea Research Part II: Topical Studies in Oceanography*.

Furlan, E.M. *et al*., 2016. A framework for estimating the sensitivity of eDNA surveys. *Molecular Ecology Resources*, 16(3).

Foster, Z.S.L. *et al*., 2017. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data T. Poisot, ed. *PLOS Computational Biology*, 13(2), p.e1005404.

Froenicke, L., 2017. Update on Barcode Mis-Assignment Issue | DNA Technologies Core.

Frøslev, T.G., Kjøller, R., Bruun, H.H., Ejrnæs, R., Brunbjerg, A.K., Pietroni, C. and Hansen, A.J., 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature communications*, 8(1), p.1188.

Furlan, E.M. & Gleeson, D., 2017. Improving reliability in environmental DNA detection surveys through enhanced quality control. *Marine and Freshwater Research*, 68(2), p.388.

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, pp.1–3.

Greer B, Maul R, Campbell K and Elliott CT. 2017. Detection of freshwater cyanotoxins and measurement of masked microcystins in tilapia from Southeast Asian aquaculture farms. Analytical and Bioanalytical Chemistry: pp.1-13.

Hajibabaei, M. *et al*., 2016. A new way to contemplate Darwin's tangled bank: how DNA barcodes are reconnecting biodiversity science and biomonitoring. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 371(1702).

Hosler, D.M., 2017. Management in Practice CORRECTED PROOF Where is the body? Dreissenid mussels, raw water testing, and the real value of environmental DNA Introduction to Reclamation Project background and goal. *Management of Biological Invasions*, 8.

Hunter, M.E. *et al*., 2017. Detection limits of quantitative and digital PCR assays and their influence in presence-absence surveys of environmental DNA. *Molecular Ecology Resources*, 17(2), pp.221–229.

Jerde, C.L. & Mahon, A.R., 2015. Improving confidence in environmental DNA species detection. *Molecular Ecology Resources*, 15(3).

Jiang, L. & Yang, Y., 2017. Visualization of international environmental DNA research. *CURRENT SCIENCE*, 112(8).

Kah-Wai, K. and Ali, A.B., 2000, February. Chenderoh Reservoir, Malaysia: Fish community and artisanal fishery of a small mesotrophic tropical reservoir. In ACIAR PROCEEDINGS (pp. 167-178). ACIAR; 1998.

Kelly, R.P., 2016. Making environmental DNA count. *Molecular Ecology Resources*, 16(1).

Klobucar, S.L., Rodgers, T.W. & Budy, P., 2017. At the forefront: evidence of the applicability of using environmental DNA to quantify the abundance of fish populations in natural lentic waters with additional sampling considerations. *Canadian Journal of Fisheries and Aquatic Sciences*, p.cjfas-2017-0114.

Kottelat, M. and Lim, K.K., fish species are recorded from the inland waters of Sarawak and Brunei Darussalam. 25 species are recorded for the first time from this area. Lahoz-Monfort, J.J., Guillera-Arroita, G. & Tingley, R., 2016. Statistical approaches to account for false-positive errors in environmental DNA samples. *Molecular Ecology Resources*, 16(3).

Lahoz-Monfort, J.J., Guillera- Arroita, G. and Tingley, R., 2016. Statistical approaches to account for false- positive errors in environmental DNA samples. *Molecular Ecology Resources*, 16(3), pp.673-685.

Leray, M., Knowlton, N. 2017. Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. *PeerJ* 5, e3006.

Liu, S. *et al*., 2016. Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. *Molecular Ecology Resources*, 16(2).

Laroche, O. *et al*., 2017. Metabarcoding monitoring analysis: the pros and cons of using co-extracted environmental DNA and RNA data to assess offshore oil production impacts on benthic communities. *PeerJ*, 5, p.e3347.

MacDonald, A.J. & Sarre, S.D., 2016. A framework for developing and validating taxon-specific primers for specimen identification from environmental DNA. *Molecular Ecology Resources*.

Mächler, E., Deiner, K., Steinmann P., Altermatt, F. 2014. Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. *Freshwater Science* 33:1174–1183.

Majaneva, M., Diserud, O.H., Eagle, S.H., Boström, E., Hajibabaei, M. and Ekrem, T., 2018. Environmental DNA filtration techniques affect recovered biodiversity. *Scientific reports*, 8(1), p.4682.

Mak, S.S.T. *et al*., 2017. Comparative performance of the BGISEQ-500 versus Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. *GigaScience*.

Meyer, M. & Kircher, M., 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Laboratory Press*

McKee, A.M., Spear, S.F. & Pierson, T.W., 2015. The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biological Conservation*, 183.

Miya, M. *et al*., 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society Open Science*, 2(7), p.150088.

O'Donnell, J.L. *et al*., 2016. Indexed PCR Primers Induce Template-Specific Bias in Large-Scale DNA Sequencing Studies A. R. Mahon, ed. *PLOS ONE*, 11(3), p.e0148698.

Olds, B.P. *et al*., 2016. Estimating species richness using environmental DNA. *Ecology and Evolution*, 6(12), pp.4214–4226.

Pfleger, M.O. *et al*., 2016. Saving the doomed: Using eDNA to aid in detection of rare sturgeon for conservation (Acipenseridae). *Global Ecology and Conservation*, 8.

Piñol, J. *et al*., 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, 15(4).

Rees, H.C. *et al*., 2015. Applications and limitations of measuring environmental DNA as indicators of the presence of aquatic animals. *Journal of Applied Ecology*, 52(4).

Roussel, J.M. *et al*., 2015. The downside of eDNA as a survey tool in water bodies. *Journal of Applied Ecology*, 52(4).

Sato, H., Sogo, Y., Doi, H. and Yamanaka, H., 2017. Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. *Scientific reports*, *7*(1), p.14860.

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(6), pp.1289–1303.

Schultz, M.T. & Lance, R.F., 2015. Modeling the sensitivity of field surveys for detection of environmental DNA (eDNA). *PLoS ONE*, 10(10).

Shelton, A.O. *et al*., 2016. A framework for inferring biological communities from environmental DNA. *Ecological Applications*, 26(6).

Sigsgaard, E.E. *et al*., 2015. Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. *Biological Conservation*, 183.

Sigsgaard, E.E. *et al.*, 2017. Seawater environmental DNA reflects seasonality of a coastal fish community. *Marine Biology*, 164(6), p.128.

Serrao, N.R., Reid, S.M. & Wilson, C.C., 2017. Establishing detection thresholds for environmental DNA using receiver operator characteristic (ROC) curves. *Conservation Genetics Resources*, pp.1–8.

Sinha, R. *et al*., 2017. Index Switching Causes "Spreading-Of-Signal" Among Multiplexed Samples In Illumina HiSeq 4000 DNA Sequencing. *bioRxiv*.

Sokal, R.R., 1963. The principles and practice of numerical taxonomy. *Taxon*, pp.190-199.

Spens, J. *et al.*, 2017. Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods in Ecology and Evolution*, 8(5), pp.635-645.

Sneath, P.H. and Sokal, R.R., 1973. *Numerical taxonomy. The principles and practice of numerical classification.* San Francisco, W.H. Freeman and Company. ISBN 0716706970.

Stat, M., Huggett, M.J., Bernasconi, R., DiBattista, J.D., Berry, T.E., Newman, S.J., Harvey, E.S. and Bunce, M., 2017. Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, 7(1), p.12240.

Sterivex^{TM} (2013) User Guide SterivexTM-GP Sterile Vented Filter Unit, 0.22 µm Single Use Only. EMD Millipore Corporation, Billerica, MA, USA.

Stewart, K. *et al*., 2017. Using environmental DNA to assess population-wide spatiotemporal reserve use. *Conservation Biology*.

Stone R. 2011. Mayhem on the Mekong. Science 333: 814.

Thomas, A.C. *et al*., 2016. Quantitative DNA metabarcoding: Improved estimates of species proportional biomass using correction factors derived from control material. *Molecular Ecology Resources*, 16(3).

Thomsen, P., Kielgast, J.O.S., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., Orlando, L. and Willerslev, E., 2012. Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular ecology*, *21*(11), pp.2565-2573.

Thomsen, P.F. *et al.*, 2016. Environmental DNA from Seawater Samples Correlate with Trawl Catches of Subarctic, Deepwater Fishes A. R. Mahon, ed. *PLoS ONE*, 11(11), p.e0165252.

Turner, C.R. *et al.*, 2014b. Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution*, 5(7).

Uchii, K., Doi, H. & Minamoto, T., 2016. A novel environmental DNA approach to quantify the cryptic invasion of non-native genotypes. *Molecular Ecology Resources*, 16(2).

Valdez-Moreno, M., Ivanova, N.V., Elias-Gutierrez, M., Pedersen, S.L., Bessonov, K. and Hebert, P.D., 2018. Using eDNA to biomonitor the fish community in a tropical oligotrophic lake. bioRxiv, p.375089.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F. and Gaboriaud, C., 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, 25(4), pp.929-942.

Vences, M. *et al*., 2016. Freshwater vertebrate metabarcoding on Illumina platforms using double-indexed primers of the mitochondrial 16S rRNA gene. *Conservation Genetics Resources*, 8(3).

Welcomme, R.L., Baird, I.G., Dudgeon, D., Halls, A., Lamberts, D. and Mustafa, M.G., 2016. Fisheries of the rivers of Southeast Asia. Freshwater *Fisheries Ecology*, pp.363-376.

Wilcox, T.M. *et al*., 2015. The dual challenges of generality and specificity when developing environmental DNA markers for species and subspecies of Oncorhynchus. *PLoS ONE*, 10(11).

Wilcox, T.M. *et al*., 2016. Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char Salvelinus fontinalis. *Biological Conservation*, 194.

Williams, K.E. *et al*., 2017. Clearing muddied waters: Capture of environmental DNA from turbid waters H. Doi, ed. *PLOS ONE*, 12(7), p.e0179282.

Wilson, C.C., Wozney, K.M. & Smith, C.M., 2015. Recognizing false positives: Synthetic oligonucleotide controls for environmental DNA surveillance. *Methods in Ecology and Evolution*.

Willoughby, J.R. *et al*., 2016. The importance of including imperfect detection models in eDNA experimental design. *Molecular Ecology Resources*, 16(4).

Winemiller KO, *et al*. 2016. Balancing hydropower and biodiversity in the Amazon, Congo, and Mekong. *Science* 351(6269): 128–129.

Yamanaka, H. & Minamoto, T., 2016. The use of environmental DNA of fishes as an efficient method of determining habitat connectivity. *Ecological Indicators*, 62.

Appendix 1

Environmental DNA for wildlife biology and biodiversity monitoring

Bohmann, K., **Evans, A**., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Douglas, W.Y. and De Bruyn, M., 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, 29(6), pp.358-367

Review

CellPress

Environmental DNA for wildlife biology and biodiversity monitoring

Kristine Bohmann[1,2*], Alice Evans[3*], M. Thomas P. Gilbert[1,4], Gary R. Carvalho[3], Simon Creer[3], Michael Knapp[3], Douglas W. Yu[5,6], and Mark de Bruyn[3]

[1] Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5–7, 1350 Copenhagen K, Denmark
[2] School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK
[3] Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Deiniol Road, Bangor University, Bangor LL57 2UW, UK
[4] Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia 6845, Australia
[5] State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 32 Jiaochang East Road, Kunming, Yunnan 650223, China
[6] School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, Norfolk NR4 7TJ, UK

Extraction and identification of DNA from an environmental sample has proven noteworthy recently in detecting and monitoring not only common species, but also those that are endangered, invasive, or elusive. Particular attributes of so-called environmental DNA (eDNA) analysis render it a potent tool for elucidating mechanistic insights in ecological and evolutionary processes. Foremost among these is an improved ability to explore ecosystem-level processes, the generation of quantitative indices for analyses of species, community diversity, and dynamics, and novel opportunities through the use of time-serial samples and unprecedented sensitivity for detecting rare or difficult-to-sample taxa. Although technical challenges remain, here we examine the current frontiers of eDNA, outline key aspects requiring improvement, and suggest future developments and innovations for research.

From sampling organisms to sampling environments

In 1966, the writers of *Star Trek* introduced intergalactic battles, alien invaders, and technology beyond the realm of reality. When the handheld Tricorder was used by Spock to test unexplored habitats, little did the writers know that the sci-fi technology to analyse an environment and its living components from a small sample would become a reality in just 50 Earth years. Free DNA molecules are ubiquitous, released from skin, mucous, saliva, sperm, secretions, eggs, faeces, urine, blood, root, leaves, fruit, pollen, and rotting bodies and are collectively referred to as eDNA (see Glossary) [1]. Any given environmental sample will contain myriad eDNA and the information contained therein is now accessible owing to advances in sample preparation and sequencing technology. Today, science fiction is becoming reality as a growing number of biologists are using eDNA for species detection and biomonitoring, circumventing, or at least alleviating, the need to sight or sample living organisms. Such approaches are also accelerating the rate of discovery, because no *a priori* information about the likely species found in a particular environment is required to identify those species. Those working on invasive species, community and ecosystem processes underpinning biodiversity and functional diversity, and wildlife and conservation biology are likely to benefit the most through adoption of eDNA techniques. Current barriers to the use of eDNA include the requirement for extensive training in molecular biology and

Corresponding authors: Gilbert, M.T.P. (mtpgilbert@gmail.com); de Bruyn, M. (markus.debruyn@gmail.com).

Keywords: biodiversity; monitoring; wildlife; environmental DNA; metabarcoding; metagenomics; second-generation sequencing.

*These authors contributed equally to this work.

0169-5347/
 http://dx.doi.org/10.1016/j.tree.2014.04.003

Glossary

Amplicon: a fragment of DNA or RNA created by replication events or amplification, either naturally or artificially, through, for example, PCR.
Ancient DNA (aDNA): DNA extracted from specimens that have not been intentionally preserved for genetic analysis. Such samples are typically low quality and can include specimens from museum collections, archaeological finds, and subfossil remains of tissues or other DNA-containing sources (e.g., coprolites, hair).
Blocking primer: an oligonucleotide used to bind to DNA and overlap the primer-binding sites, so that amplification of the undesired species is prevented.
Chimera: sequences that arise during amplification combining DNA fragments from two or more individuals.
Environmental DNA (eDNA): trace DNA in samples such as water, soil, or faeces. eDNA is a mixture of potentially degraded DNA from many different organisms. It is important to note that this definition remains controversial due to the sampling of whole microorganisms that might appear in an environmental sample. Although metagenomic microbial studies might use environmental sampling, they cannot always be defined as true eDNA studies because some methods first isolate microorganisms from the environment before extracting DNA.
Metagenomics: sequencing of the total DNA extracted from a sample containing many different organisms.
Operational taxonomic unit (OTU): the taxonomic level of sampling defined by the researcher in a study; for example, individuals, populations, species, genera, or strains. OTUs are generated by comparing sequences against each other to form a distance matrix, followed by clustering groups of sequences with a specified amount of variability allowed within each OTU (e.g., [67]).
Second-generation sequencing: sequencing technologies such as the Roche GS series, Illumina Genome Analyser series, and IonTorrent series that parallelise the sequencing process, producing thousands to billions of DNA sequences in single sequencing runs.

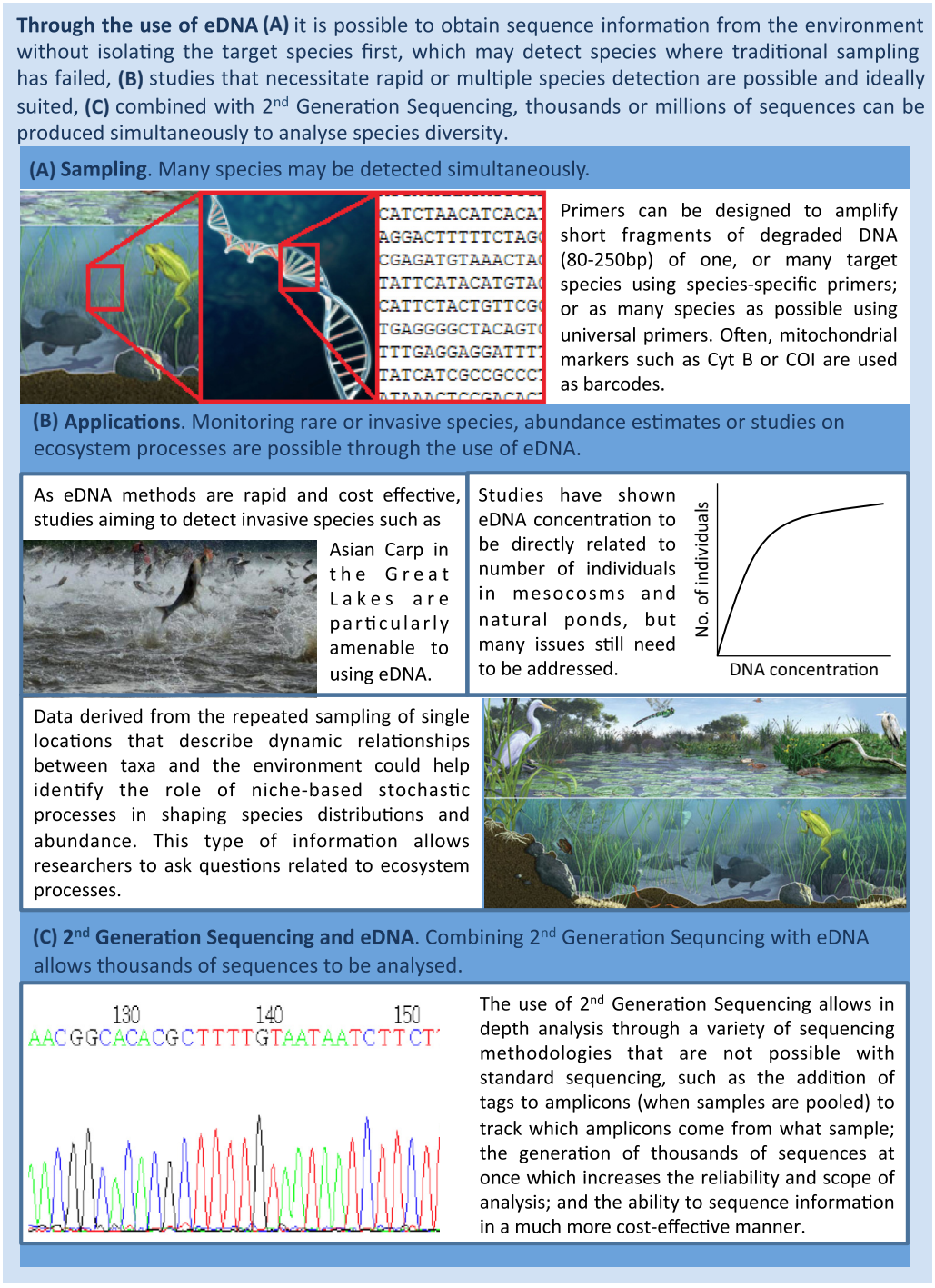


Figure 1. Summary of **(A)** the concept of environmental DNA (eDNA), **(B)** promising applications of eDNA, and **(C)** the advantages of combining eDNA with second-generation sequencing.

the subsequent genetic data analysis; however, the rapid emergence of commercial companies specialising in eDNA [e.g., SpyGen (http://www.spygen.fr/en/)] provides a way around this analytical bottleneck.

As the technologies have improved, the ability to detect trace quantities of eDNA and the breadth of environments more readily accessible to researchers have increased dramatically (Figure 1). Although the field of metagenomics (the study of many genomes) and metagenetics (the study of many genes) has until recently been considered applicable only to microorganisms, the idea of metagenetics in a macrobial sense is being applied to samples of eDNA in trace amounts left behind in the environment by organisms which are no longer present, as opposed to whole microorganisms that have been used in the latter fields. Such an approach facilitates community eDNA analysis [2] simultaneously from across the kingdoms of life, including, for example, plants, animals, fungi, and bacteria [3] (examples of which are shown in Table 1). In addition, eDNA offers researchers a glimpse of the DNA from elusive and endangered species,

Table 1. Examples of the wide range of eDNA applications

Sample	Application	Studies of importance	Refs
Applications with potential for conservation biology and policy-making decisions			
Blood meal	Species detection	DNA of rare mammals such as the elusive Truong Son muntjac (*Muntiacus truongsonensis*) identified in leeches collected in Vietnam	[58]
Faeces	Population genetics	Highly fragmented and isolated populations of giant panda (*Ailuropoda melanoleuca*) were analysed and landscape genetic patterns, divergence time, and population structure identified	[68]
Honey	Species detection	Plant and insect DNA identified in just 1 ml of honey	[69]
Seawater	Species detection	Harbour porpoise (*Phocoena phocoena*) and long-finned pilot whale (*Globicephala melas*) detected in the western Baltic	[30]
Snow	Species detection	Wolf (*Canis lupus*) DNA isolated from blood spots in the Italian Alps and Arctic fox (*Alopex lagopus*) DNA isolated from footprints	[70,71]
Soil	Species detection	Vertebrate mitochondrial DNA (mtDNA) identified in soil samples collected in a zoological garden and a safari park matched to the elephant and tiger inhabitants, respectively	[29]
Applications with potential for ecology (including palaeo- and macroecology)			
Cave sediments	Reconstructing past flora and fauna	Extinct biota identified from cave sediment in New Zealand, revealing two species of ratite moa and 29 species of plants from the prehuman era	[42]
Fresh water	Species detection and biomass estimation	Diversity of rare and threatened freshwater fish, amphibians, mammals, insects, and crustaceans was quantified in eDNA from small water samples collected in lakes, ponds, and streams	[28]
Ice cores	Reconstructing past flora and fauna	Plant and insect diversity from the past million years was catalogued from deep ice cores in Greenland	[72]
Nunatak sediments	Reconstructing past flora and fauna	Reconstruction of vegetation from the end of the Holocene Thermal Maximum [5528 $\pm$ 75 calibrated years before present (BP)] from bedrock protruding through ice sheets (nunatak sediments)	[43]
Permafrost	Reconstructing past flora and fauna, habitat conservation	Fungal, bryophyte, enchytraeid, beetle, and bird DNA identified in frozen sediment of late-Pleistocene age (circa 16 000–50 000 years BP)	[73, reviewed in 74]
Saliva/twigs	Species detection	DNA in saliva on browsed twigs identified browsing moose (*Alces alces*), red deer (*Cervus elaphus*), and roe deer (*Capreolus capreolus*), amplifying in some samples up to 24 weeks after the browsing event	[75]
Applications with potential for the understanding of ecosystems			
Air	Invasive-species detection	The presence of genetically modified organisms was detected from samples of air containing low levels of pollen	[76]
Fresh water	Wildlife-disease detection	Detecting the chytrid fungus *Batrachochytrium dendrobatidis*, which is likely to be a primary cause of amphibian population declines, in water samples	[77]
Fresh water	Invasive-species detection	The American Bullfrog (*Lithobates catesbeianus*) was successfully identified, showing that early detection of invasive species at low densities is possible and has implications for management	[44]

undetected invasive species, and species in habitats where they were previously unrecorded due to difficulty in locating such species or their active avoidance of conventional sampling methods. To date, in addition to proof of principle, eDNA studies have predominantly focused on species identification, as well as the detection of pathogenic, endangered, invasive, genetically modified, and game species and the reconstruction of diets and ancient communities (Table 1).

There is now sufficient evidence that natural processes continuously deposit DNA into the environment in ways that make it possible to reconstruct ecological and evolutionary processes from easy-to-collect samples. Open questions include how accurate, unbiased, and detailed the eDNA record is and how best to extract and analyse the genetic information with the technologies currently available today – points of particular relevance because DNA degrades rapidly once exposed to oxygen, light, heat, DNases, or water [4]. Like the related study of ancient DNA (aDNA) (e.g., [5]), eDNA approaches require rigorous standards and controls, without which the information obtained might not only be noisy, but outright misleading.

A substantial eDNA literature now exists, which we draw on below to ask what will and could be achieved through the use of eDNA and how it will and could change what we understand about species and ecosystems. To do so, we discuss how eDNA approaches can be used to examine timely questions in ecology and evolution and consider how such insights might contribute to advances in these fields. The recent surge in eDNA studies, facilitated to a large extent by recent technological advances in affordable high-throughput sequencing, demands a critique of this emerging fields' scope of application as well as its limitations, to facilitate uptake of nascent opportunities while maintaining scientific rigour. We highlight particularly promising areas of eDNA research and evaluate priorities for additional work.

Describing ecosystem-level processes

Realistic inferences and predictions about the impact of environmental change on extant biota depend increasingly on our ability to transcend boundaries among traditional biological hierarchies in the wild, extending from individuals to species, populations, and communities. The implementation of so-called ecosystem-based approaches [6],

which take a more holistic view than single-species studies, is particularly amenable to eDNA, where trophic, energetic, and terrestrial–aquatic interactions can be detected and tracked. A recent demonstration of such functional links to biodiversity [7] was among the first to link functional traits and DNA metabarcoding studies. Using community traits from metagenomic aquatic samples, significant differences were detected between the community profiles derived from the commonly used 16S rRNA gene and from functional trait sets. Traits yielded informative ecological markers by discriminating between marine ecosystems (coastal versus open ocean) and oceans (Atlantic versus Indian versus Pacific). Another recent study [8] used eDNA for a community analysis in an ecotoxicology setting. This study examined the effect of elevated levels of triclosan, a common antibiotic and antifungal agent used in many consumer goods, on benthic invertebrate communities through microcosm experiments, and observed a pronounced loss of metazoan operational taxonomic units (OTUs) due to increased levels of triclosan.

Key ecosystems underpinning plant biological production and carbon and nutrient cycling can also be readily characterised using eDNA washed from root systems [9], generating insights into the dynamics of community structure and providing an ecological framework to investigate functional links among root-associated fungi, environmental variation and ecosystem diversity, and associated services. In this context, complementary multidisciplinary approaches, such as combining eDNA with aDNA and morphological analyses of micro- and macrofossils, show particular promise for elucidating the impact of changing climates on species and communities through time [3,10–13]. Macroecology, for example, is undergoing a small revolution as studies based on environmental samples transform our understanding of microorganismal abundance, range size, and species richness (e.g., [14–16]). Such insights provide a major impetus for understanding the distribution and drivers of diversity on our planet, from megafauna to viruses, particularly in regions that are difficult to study using more traditional methods (e.g., Antarctic lakes [17], deep-sea anoxic basins [18]).

One of the main advantages of eDNA approaches to understanding ecosystems is the relative ease with which eDNA samples can be collected, which enables researchers to analyse the dynamics of community diversity through time. Rather than looking at static snapshots that are limited by the difficulty of observation, researchers can now easily sample species in an area as often as geography permits, creating what could be imagined as a 'stop-motion eDNA video'. Moreover, data derived from repeated sampling of single locations could help identify the role of niche-based and stochastic processes in shaping species distributions and abundance [19].

Using eDNA to estimate relative abundance

A major opportunity provided by quantitative analysis of eDNA is to move beyond measures of the presence–absence of a species to its relative abundance in natural systems [20,21]. Such abundance estimates are, however, not straightforward. Although presence–absence measures can provide useful indicators of biological diversity, they are often insufficient to link biological diversity to ecosystem functioning [22]. Similarly, the ability to detect rare or endangered species with confidence is of clear conservation value, but mere presence does not necessarily indicate recruitment or persistence in a given habitat. Rapid measures of abundance or biomass across time and space would be more informative and, importantly, can reveal seasonal shifts in factors such as microhabitat use for feeding and/or reproduction or refuge use, as well as impacts of predation and competition. Approaches to date to estimate abundance using eDNA include [20], which used eDNA to detect Asian carp, and repeated sampling to generate an abundance index thereof (see also [23–25]); [26] showed that rank abundance of recovered fish eDNA sequences correlated with the abundance of the corresponding species' biomass in a large mesocosm; whereas [27] extended this and used occupancy models to correct for the fact that even eDNA has a less-than-perfect detection probability. An additional way to estimate abundance estimation is to base it on DNA concentrations.

The opportunity to estimate abundance based on concentrations of eDNA relies in part on the assumption that the release of eDNA from faeces, secretions, or tissues is correlated with the abundance or standing biomass of the respective individuals. Although such correlations have been demonstrated in a few studies (e.g., [28,29]), there are three core challenges that must be overcome before informative relative abundance data can be generated. First, robust information on the persistence of eDNA in the wild from a broad range of climates and habitats is necessary. It is well established that eDNA decay rates vary considerably under different environmental conditions [30–32], which will result in biased estimates of abundance. Second, our understanding of how environmental factors, including digestive systems for faecal matter-based studies, affect eDNA concentrations needs to be improved [33–36]. Finally, the assumption needs to be tested that eDNA sequence copy numbers accurately reflect the original composition of DNA in an environmental sample [37] and are not altered somewhere along the analytical pipeline (Box 1).

Water sampling illustrates the complexity of interpreting eDNA-based studies. Detection probability is likely to be dependent on the interplay between the density of target species, the amount of DNA released via excretion, and variation in rates of dilution and diffusion depending on the environment, temperature, microbial communities, and the rate of DNA degradation, to name but a few of the variables. In the studies performed to date (e.g., [25,28,32]), waterborne eDNA appears to yield near-real-time, local, and reliable-but-noisy estimates of species frequencies, although DNA concentration may fall to sub-detectable levels once organisms are removed from the environment over relatively short time spans (around 2 weeks in Northern European artificial ponds [28]). By contrast, in soil or lake sediments, detectable traces of plant and animal eDNA persist for centuries or millennia (e.g., [33,38–41]) or even tens to hundreds of millennia when frozen (e.g., [10,41–43]). Comprehensive replicated sampling surveys are required to evaluate eDNA abundance and dynamics across a range of species and study sites.

Box 1. Improving eDNA data recovery in the laboratory

Recent years have seen rapid improvements in sequencing technologies and we are only beginning to see the associated opportunities for eDNA research. However, continued improvements to current eDNA protocols are conceivable for all aspects of laboratory work.

Sequencing library preparation

Future eDNA studies are likely to take an increasingly metagenomic approach. Instead of PCR enriching a relatively small number of markers before sequencing, the eDNA extract will be sequenced in its entirety. If, however, PCR is avoided completely, libraries have to be prepared directly from potentially highly degraded eDNA. Most existing library preparation protocols are optimised for high-quality DNA and are inefficient for highly degraded DNA [78–80]. To overcome this limitation, eDNA methods can benefit from developments in the field of aDNA, which routinely produces potentially relevant protocols in this regard (e.g., [79]).

Target enrichment

Until the sequence output of second-generation sequencing platforms becomes sufficient to avoid informative marker targeting, enrichment methods are needed. Although PCR represents the basic option, hybridisation-based sequence capture might offer an alternative [81]. With an ability to target short molecules, under relatively permissive levels of mismatch [82] such methods might bypass major disadvantages of PCR enrichment.

Blocking of undesired molecules

A further approach to increase the percentage of informative markers is to prevent non-target molecules from being enriched and sequenced by sequestering them with blocking oligonucleotides (e.g., [83]). The approach has so far mostly been used to exclude a relatively small set of contaminating molecules from being sequenced. However, as the amount of eDNA sequence data increases, it is conceivable that 'blocking libraries' for common environmental contaminants will be created. For example, blocking GC-rich molecules can reduce the amount of bacterial DNA sequenced in a library.

Direct shotgun sequencing

The power of Illumina-based direct shotgun sequencing of bulk insect samples was recently demonstrated [84], with subsequent informatics recovery of informative markers from the output. By avoiding the biases introduced by all target-enrichment strategies, as sequencing costs drop and outputs increase, we might for the first time obtain directly quantifiable data representing the unbiased components of an eDNA extract. With the arrival of third-generation single-molecule sequencers (e.g., Pacific Biosciences [85], Oxford Nanopore GridION™ and MinION™ [86]) that remove the need for amplification during library building, these benefits will increase yet further.

The potential to use eDNA sequencing as a high-throughput means of obtaining measures of abundance across large scales and many taxa simultaneously offers the promise of detecting cooperative and competitive relationships through robust tests of co-occurrence. Within the next 3–5 years, a coordinated global network of eDNA surveillance and monitoring activities can be envisioned as proof of principle is established across a range of environments and their resident taxa, moving eDNA from an emerging field to one at the forefront of biodiversity science. The applicability of such data would provide a potential framework for global ecosystem network prediction and enable the development of ecosystem-wide dynamic models [22]. Such analyses will, for example, allow exploration of long-standing issues relating to the nature and dynamics of shifts in community assembly (e.g., [3,10,41–43]).

eDNA in applied conservation biology

One of the most attractive facets of eDNA is its potential as a rapid and cost-effective tool for applied conservation biology, including early detection of invasive species and monitoring of otherwise difficult-to-detect species. The use of eDNA as an early-warning system for the detection of invasive species [20,44–46] and pathogens [47] at low density, at any life stage or season, and through *ad hoc* sampling of substrates as diverse as ship ballast water, aquaculture transits, or habitats at high risk can alert regulatory authorities before the establishment of alien species. Indeed, the method has already demonstrated particular promise. The US Fish and Wildlife Service, for example, have implemented an eDNA-based approach to monitor invasive Asian carp in the Midwest, USA (Figure 2A), providing a labour- and cost-effective alternative to traditional large-scale sampling methods such as electrofishing and/or manual netting [20]. Uptake of eDNA methodologies into biomonitoring of invasive species for fisheries appears to be increasing, with events such as the American Fisheries Society symposium in September 2013 entitled 'Environmental DNA (eDNA) Analysis – a New Genetic Tool for Monitoring, Managing, and Conserving Fishery Resources and Aquatic Habitat', which covered the topics of Asian carp in the Great Lakes, the invasive New Zealand mud snail, and the invasive African jewel fish (https://afs.confex.com/afs/2013/webprogram/Session2539.html).

Despite the promise of using eDNA as an early-warning system, eliminating false positives remains a major challenge (see Box 2 for an extended discussion). The mere presence of eDNA does not necessarily indicate the presence of the relevant organism, due to the potential for eDNA dispersal (in particular for air- or waterborne eDNA) or contamination. Where there is the potential for high connectivity, such as in aquatic systems, this challenge may be tempered if the study design incorporates risk assessment of target eDNA emanating from sources such as sewage and wastewater, bilge water discharge, excrement from predatory fish or waterfowl, dead fish carried on barges and boats from elsewhere, or even carry-over from PCR and sequencing chemistries. For example, [48] shows that invertebrate eDNA can travel up to, and potentially further than, 12 km along river systems. In short, robust control of false positives to assess and control for contamination are critical in eDNA analyses, as is the case for aDNA studies (e.g., [5]).

An extension to the use of eDNA in conservation biology is its use in species monitoring through diet analyses (e.g., [49,50]). Traditionally, diet analyses were performed either by directly observing what an animal ate or by collecting its faeces and examining prey fragments under a microscope. These results were then used in ecological studies of, for example, predator ecology, interspecific competition, or niche partitioning. For some animals, however, these approaches are unfeasible, as is the case with insectivorous bats, which prey aerially in the dark and masticate or void the larger prey fragments. eDNA has provided an alternative or complementary approach and metabarcoding, in which second-generation sequencing is performed on amplicons originating from faecal or other bodily extracts amplified with tagged universal primers [51] (Figure 2C), has made it more efficient and cost-effective to obtain diet information on a large scale (e.g., [34,52–55]; reviewed in [56,57]).

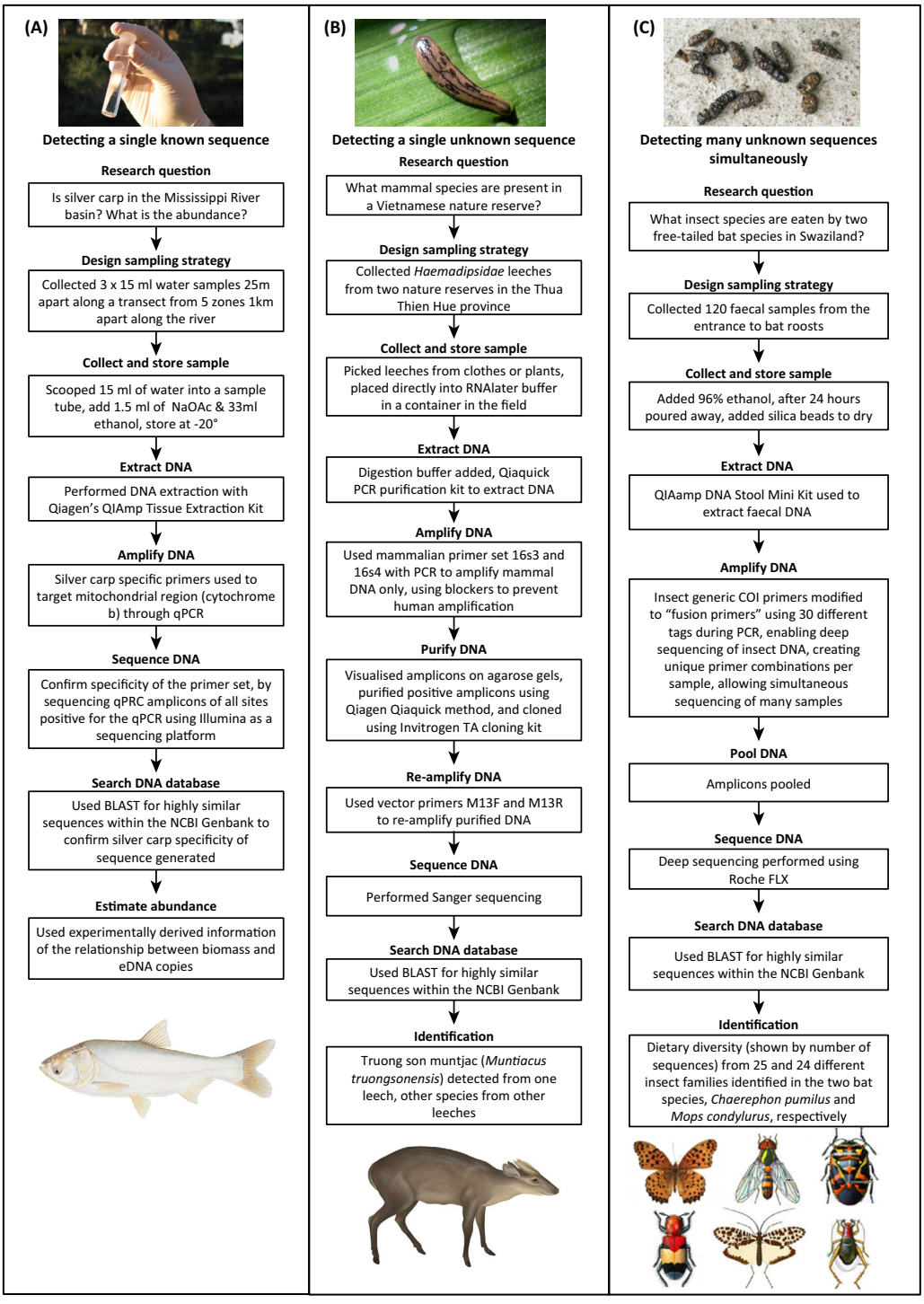


Figure 2. Exemplary environmental DNA (eDNA) case studies illustrating three research questions and the experimental procedures followed. eDNA studies can be designed in various ways to address the research question. **(A)** Detection and abundance estimation of invasive Asian carp in a water sample [20,87]. **(B)** Detection of mammal species in leech blood meals [58]. **(C)** Detection of insect prey in bat faeces [34]. Each example follows a general framework (in bold) and a specific procedure (in boxes).

Because predators or blood-sucking insects feed on biodiversity, collecting either faecal material or the insect itself for molecular diet analysis can identify rare or cryptic species that traditional monitoring methods such as camera traps might miss. Recent studies on this include stomach-content analyses of parasitic invertebrates such as leeches [58] (Figure 2B), carrion flies [59], mosquitoes [60], and ticks [61] to reveal their vertebrate hosts. In one case, Vietnamese terrestrial leeches of the genus *Haemadipsa* revealed the presence of an endemic rabbit species that had not been detected despite monitoring the site for several thousand nights with camera traps [58]. Leeches are currently being used to search for the highly endangered saola antelope in Vietnam and Laos [Saola

Box 2. Sources of uncertainty from eDNA and how they can be overcome

Source 1. False positives (type I error: eDNA detected where target species is not present) resulting from false detection of eDNA from other sources, such as tributaries into a major river, ballast water discharge, sewage and waste water, excrement from animals that prey on the target species, dead target species carried on boats, or unsterilised equipment (see [20,32,88]).

Solution 1. To ensure false positives do not occur via contamination between samples when using the same equipment, equipment must be sterilised thoroughly or, preferably, not reused [20]. Quality control to avoid false positives should be implemented in the sampling strategy; for example, blank samples can be taken into the field to ensure contamination does not occur in the transport phase [20] and samples can be taken from adjacent areas where target species are known not to occur [20]. Sampling design should incorporate a risk assessment of target and non-target eDNA.

Source 2. False positives resulting from PCR primers and probes that do not have a high enough level of specificity, allowing the amplification of 'lookalike' non-target DNA [32,45,88].

Solution 2. In silico testing of species-specific DNA-based probes and primers, such as comparing sequences with the Basic Local Alignment Search Tool (BLAST), or using ecoPCR software, as well as *in vitro* testing of probes and primers against target tissue-derived DNA [32,88]; genetic distances should also be reported [20].

Source 3. False negatives (type II error: eDNA not detected where target species is present) resulting from insufficient sensitivity or failure of methods to perform as expected [88].

Solution 3. Rigorous testing of primers against target species' DNA must be undertaken to ensure successful amplification, as well as optimising protocols to be confident of species detection before sample collection begins.

Source 4. The inability of eDNA to distinguish between live or dead organisms [88], including digested or faecal remains of target organisms derived from their predators (e.g., birds preying on fish).

Solution 4. Repeated temporal sampling of the same area will alleviate this problem to some extent. Because dead bodies, predators' faecal matter, or other introduced sources of DNA decompose and degrade over time, a species that is permanently present in an environment will still be detected after the introduced contaminants have degraded beyond the point of DNA amplification. The study's risk assessment should include any visually observed dead organisms.

Working Group (2013) *Conservation Through Collaboration: Proceedings of the 3rd Meeting of the Saola Working Group 2013* (http://www.savethesaola.org)].

Advantages of eDNA as an assessment tool

Although advances in technology can themselves propel new conceptual insights, uptake will depend crucially on the cost-effectiveness of any new tools and the ease and efficacy of the approach. It is worth noting that, as with the introduction of DNA barcoding *sensu stricto* [62], which aimed to complement the Linnaean system of taxonomy, eDNA will most likely exert a pervasive impact through its integration with existing approaches rather than necessarily replacing them. A study from 2012 [30] demonstrates the advantage of this combined approach. By evaluating the use of eDNA in detecting marine mammals, it was shown that conventional static acoustic monitoring devices that recognise echolocation were more effective in detecting the harbor porpoise (*Phocoena phocoena*) in natural environments; however, eDNA detected the rare long-finned pilot whale (*Globicephala melas*), demonstrating how eDNA and conventional sampling can work together.

Recent work on eDNA from water samples (e.g., endangered hellbender salamanders [*Cryptobranchus a. alleganiensis*] [63]) demonstrates the benefits of eDNA analysis, which not only is less labour intensive but, importantly, is noninvasive, thereby minimising disruption to already fragile microhabitats and reducing disease transfer and stress to target species. Filtering of water samples in this case enabled the reliable detection of target eDNA even where specimens occurred at low frequencies (as also shown in [28,30,31]). In the case of the hellbender salamander, the greatest saving was in person-hours; whereas, typically, large teams are required for traditional sampling by rock lifting, a single researcher can collect and filter water. Another example in this context examined direct comparisons between eDNA and traditional estimates based on auditory and visual inspection of the invasive American bullfrog *Rana catesbeiana* [44]. Findings revealed a higher efficiency of the former in both sensitivity and sampling effort.

Various cost-effective and simple protocols can be employed to enhance effectiveness. With a diverse array of sampling (e.g., water/soil volume), concentrating (e.g., precipitation versus filters), DNA extraction (e.g., kits and protocols), primer optimisation, and PCR protocols (e.g., efficacy of quantitative PCR [qPCR] [64]) available, it is of high priority to compare their efficacy and application under a range of biological and abiotic conditions [65]. Protocols and sampling kits can be developed to enable citizen-science approaches, such as that proposed by the Freshwater Habitats Trust and partners (Spygen, ARC and University of Kent) in the UK. In 2013 this group undertook an extensive trial of the eDNA approach to test for the presence and abundance of the endangered great crested newt (*Triturus cristatus*) in British freshwaters. Results were promising [93] and suggest that community engagement with eDNA sampling is feasible; however, they, along with the stakeholders, methodological developers, resource managers, and policy makers, must be made aware of the current levels of uncertainty associated with eDNA (discussed in Box 2). This is critical when eDNA methodology is being used to inform management or development decisions, such as those faced by local planning authorities responsible for enforcing environmental regulations with regard to planning developments and endangered species.

The future of eDNA in ecology and wildlife monitoring

It is enticing to imagine the possibilities that eDNA could open up, if advances in molecular ecology, bioinformatics, and sequencing technologies continue to accelerate. The main advantages of eDNA are rooted in its autonomous nature; with a reduced need for human taxonomists, ecologists, or biologists, sampling can access inhospitable environments, target elusive species, and provide a vast reduction in labour costs. In the future, it may be possible to implement mechanical sampling of eDNA, similar to that of oil spill-sampling buoys or military sonobuoys. When combined with the technology to transmit live data such as that used by the US National Weather Service (http://earth.nullschool.net/), technology currently being developed by Oxford Nanopore Technologies to sample

Box 3. Outstanding questions

- Can we catalogue the variables that will affect eDNA half-life and can we set standards to determine whether the samples are degraded past the point of use (e.g., [32,89])?
- How do we best preserve samples for later analyses of eDNA (e.g., [90])?
- What are the dispersive properties of eDNA in various environments (e.g., [33,91]) – how readily is eDNA transported between horizons and environments (e.g., [92])?
- How can we more rapidly and cost-effectively analyse field samples? One method still in the testing phase is a mobile DNA sampler that sends results to the laboratory directly from the field (http://www.environmentalhealthnews.org/ehs/news/2013/beach-tests).
- As with the field of metagenomics, how can we more powerfully and reliably define and assign taxonomies to eDNA sequences?
- How quantitative is eDNA data – can conversion factors be meaningfully implemented to account for sampling, biomass, and amplification biases?

and analyse DNA using a handheld MinION™ device, and the current ongoing project to map the Earth's surface in 3D (http://www.bbc.co.uk/news/science-environment-16578176), it is not beyond the realm of possibility to imagine a situation where eDNA videos could be recorded in real time from automated sampling stations. Such stations could remotely relay sequence information of interest, with additional data overlaid, – including, for example, water depth, hydrological or other environmental movements, temperature, and pH – that could help identify how long eDNA had been in the environment and where it was likely to have originated from. On a smaller scale, this approach could be applied to human samplers targeting environments of interest, sampling eDNA, and remote uploading information via smartphone, creating a network of live biodiversity assessment, or the implementation of 'eDNA traps' similar to camera traps. On a larger scale, this approach could be applied to the sampling of inaccessible habitats, such as the Arctic or the deep sea, by remote samplers.

A more realisable goal in the short term is the potential for the use of eDNA in population genetics, with, for example, applications for conservation genetics and phylogeography. To date, to the best of our knowledge, such an approach has not yet been attempted. If eDNA stores sufficient population-specific information within molecular markers (e.g., mitochondrial haplotypes), it is possible that eDNA could be used directly for population genetic studies. With repeated sampling across temporal and geographical scales, this information could feed in to questions related to biogeography or palaeoecology.

The next step for eDNA

eDNA has proven its worth in detecting not only common species, but also endangered, undetected invasive, or elusive native species. As with most technological advances, limitations remain, as do many challenges that need to be overcome to move beyond mere species detection (Box 3). The potential implementation of eDNA approaches across disciplines indicates that it will be critical not only to sample, extract, and sequence eDNA in an efficient and cost-effective manner, but also to handle and analyse efficiently and reliably the typically massive data sets generated by second-generation sequencing platforms. Future eDNA studies should aim to refine and improve the processing, analysing, and organisation of what has been referred to as a 'tidal wave' of sequence information [66]. Although detailed bioinformatic considerations are beyond the scope of this review, they are crucial to consider when conducting an eDNA study. Although eDNA methods applicable to a broad range of environments and their resident taxa are currently being tried and tested, work remains to be done to ensure their reliability and repeatability (Box 1), particularly with regard to false positives and negatives (Box 2). The current evidence outlined above indicates that such effort is warranted, with exemplary eDNA studies including multiple approaches to address such uncertainties (Box 2). eDNA is on the brink of making significant contributions to our understanding of invasive species, community and ecosystem processes underpinning biodiversity and functional diversity, and wildlife and conservation biology.

Acknowledgements

The authors thank the Editor, Paul Craze, and three anonymous reviewers for helpful comments that greatly improved the manuscript. A fourth reviewer, Doug Yu, was invited to join the writing team, and agreed to do so, after contributing a detailed and extremely insightful initial review.

References

1 Taberlet, P. *et al.* (2012) Environmental DNA. *Mol. Ecol.* 21, 1789–1793
2 Porco, D. *et al.* (2010) Coupling non-destructive DNA extraction and voucher retrieval for small soft-bodied arthropods in a high-throughput context: the example of Collembola. *Mol. Ecol. Resour.* 10, 942–945
3 Willerslev, E. *et al.* (2014) Fifty thousand years of arctic vegetation and megafauna diet. *Nature* 506, 47–51
4 Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* 362, 709–715
5 Gilbert, M.T.P. *et al.* (2005) Assessing ancient DNA studies. *Trends Ecol. Evol.* 20, 541–544
6 Clarke, P. and Jupiter, S. (2010) *Principles and Practice of Ecosystem-based Management – a Guide for Conservation Practitioners in the Tropical Western Pacific*, Wildlife Conservation Society
7 Barberán, A. *et al.* (2012) Exploration of community traits as ecological markers in microbial metagenomes. *Mol. Ecol.* 21, 1909–1917
8 Chariton, A.A. *et al.* (2014) A molecular-based approach for examining responses of eukaryotes in microcosms to contaminant-spiked estuarine sediments. *Environ. Toxicol. Chem.* 33, 359–369
9 Blaalid, R. *et al.* (2012) Changes in the root-associated fungal communities along a primary succession gradient analysed by 454 pyrosequencing. *Mol. Ecol.* 21, 1897–1908
10 Jorgensen, T. *et al.* (2012) A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Mol. Ecol.* 21, 1989–2003
11 Anderson-Carpenter, L.L. *et al.* (2011) Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. *BMC Evol. Biol.* 11, 30
12 Lejzerowicz, F. *et al.* (2013) Ancient DNA complements microfossil record in deep-sea subsurface sediments. *Biol. Lett.* http://dx.doi.org/10.1098/rsbl.2013.0283
13 Der Sarkissian, C. *et al.* (2014) Shotgun microbial profiling of fossil remains. *Mol. Ecol.* 23, 1780–1798
14 Ogram, A. *et al.* (1987) The extraction and purification of microbial DNA from sediments. *J. Microbiol. Methods* 7, 57–66
15 Bik, H.M. *et al.* (2012) Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol. Evol.* 27, 233–243
16 Zinger, L. *et al.* (2011) Two decades of describing the unseen majority of aquatic microbial diversity. *Mol. Ecol.* 21, 1878–1896

17 Bissett, A. *et al.* (2005) Isolation, amplification, and identification of ancient copepod DNA from lake sediments. *Limnol. Oceanogr. Methods* 3, 533–542
18 Corinaldesi, C. *et al.* (2011) Preservation, origin and genetic imprint of extracellular DNA in permanently anoxic deep-sea sediments. *Mol. Ecol.* 20, 642–654
19 Haegeman, B. and Loreau, M. (2011) A mathematical synthesis of niche and neutral theories in community ecology. *J. Theor. Biol.* 269, 150–165
20 Jerde, C.L. *et al.* (2011) "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conserv. Lett.* 4, 150–157
21 Minamoto, T. *et al.* (2012) Surveillance of fish species composition using environmental DNA. *Limnology* 13, 193–197
22 Faust, K. and Raes, J. (2012) Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10, 538–550
23 Ficetola, G.F. *et al.* (2008) Species detection using environmental DNA from water samples. *Biol. Lett.* 4, 423–425
24 Goldberg, C.S. *et al.* (2011) Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. *PLoS ONE* 6, e22746
25 Takahara, T. *et al.* (2013) Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLoS ONE* 8, e56584
26 Kelly, R.P. *et al.* (2014) Using environmental DNA to census marine fishes in a large mesocosm. *PLoS ONE* 9, e86175
27 Schmidt, B.R. *et al.* (2013) Site occupancy models in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. *Methods Ecol. Evol.* 4, 646–653
28 Thomsen, P.F. *et al.* (2011) Monitoring endangered freshwater biodiversity using environmental DNA. *Mol. Ecol.* 21, 2565–2573
29 Andersen, K. *et al.* (2012) Meta-barcoding of "dirt" DNA from soil reflects vertebrate biodiversity. *Mol. Ecol.* 21, 1966–1979
30 Foote, A.D. *et al.* (2012) Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. *PLoS ONE* 7, e41781
31 Thomsen, P.F. *et al.* (2012) Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE* 7, e41732
32 Dejean, T. *et al.* (2011) Persistence of environmental DNA in freshwater ecosystems. *PLoS ONE* 6, e23398
33 Haile, J. *et al.* (2007) Ancient DNA chronology within sediment deposits: are paleobiological reconstructions possible and is DNA leaching a factor? *Mol. Biol. Evol.* 24, 982–989
34 Bohmann, K. *et al.* (2011) Molecular diet analysis of two African free-tailed bats (Molossidae) using high throughput sequencing. *PLoS ONE* 6, e21441
35 Deagle, B.E. *et al.* (2010) Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? *Conserv. Genet.* 11, 2039–2048
36 Murray, D.C. *et al.* (2011) DNA-based faecal dietary analysis: a comparison of qPCR and high throughput sequencing approaches. *PLoS ONE* 6, e25776
37 Hajibabaei, M. *et al.* (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE* 6, e17497
38 Yoccoz, N.G. *et al.* (2012) DNA from soil mirrors plant taxonomic and growth form diversity. *Mol. Ecol.* 21, 3647–3655
39 Hebsgaard, M.B. *et al.* (2009) The farm beneath the sand – an archaeological case study on ancient "dirt" DNA. *Antiquity* 320, 430–444
40 Parducci, L. *et al.* (2012) Glacial survival of boreal trees in northern Scandinavia. *Science* 335, 1083–1086
41 Giguet-Covex, C. *et al.* (2014) Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nat. Commun.* 5, 3211
42 Willerslev, E. *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300, 791–795
43 Jorgensen, T. *et al.* (2012) Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA meta-barcoding. *Mol. Ecol.* 21, 1980–1988
44 Dejean, T. *et al.* (2012) Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*. *J. Appl. Ecol.* 49, 953–959
45 Mahon, A.R. *et al.* (2013) Validation of eDNA surveillance sensitivity for detection of Asian carps in controlled and field experiments. *PLoS ONE* 8, e58316
46 Piaggio, A.J. *et al.* (2014) Detecting an elusive invasive species: a diagnostic PCR to detect Burmese python in Florida waters and an assessment of persistence of environmental DNA. *Mol. Ecol. Resour.* 14, 374–380
47 Aintablian, N. *et al.* (1998) Detection of *Bordetella pertussis* and respiratory syncytial virus in air samples from hospital rooms. *Infect. Control Hosp. Epidemiol.* 19, 918–923
48 Deiner, K. and Altermatt, F. (2014) Transport distance of invertebrate environmental DNA in a natural river. *PLoS ONE* 9, e88786
49 Clare, E.L. *et al.* (2009) Species on the menu of a generalist predator, the eastern red bat (*Lasiurus borealis*): using a molecular approach to detect arthropod prey. *Mol. Ecol.* 18, 2532–2542
50 Razgour, O. *et al.* (2011) High-throughput sequencing offers insight into mechanisms of resource partitioning in cryptic bat species. *Ecol. Evol.* 1, 556–570
51 Binladen, J. *et al.* (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* 2, e197
52 Deagle, B.E. *et al.* (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Mol. Ecol.* 18, 2022–2038
53 Pegard, A. *et al.* (2009) Universal DNA-based methods for assessing the diet of grazing livestock and wildlife from feces. *J. Agric. Food Chem.* 57, 5700–5706
54 Rasmussen, M. *et al.* (2009) Response to comment by Goldberg *et al.* on "DNA from pre-Clovis human coprolites in Oregon North America". *Science* 325, 148
55 Soininen, E.M. *et al.* (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Front. Zool.* 6, 16
56 Pompanon, F. *et al.* (2012) Who is eating what: diet assessment using next generation sequencing. *Mol. Ecol.* 21, 1931–1950
57 Valentini, A. *et al.* (2009) DNA barcoding for ecologists. *Trends Ecol. Evol.* 24, 110–117
58 Schnell, I.B. *et al.* (2012) Screening mammal biodiversity using DNA from leeches. *Curr. Biol.* 22, R262–R263
59 Calvignac-Spencer, S. *et al.* (2013) Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Mol. Ecol.* 22, 915–924
60 Kent, R.J. (2009) Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. *Mol. Ecol. Resour.* 9, 4–18
61 Gariepy, T.D. *et al.* (2012) Identifying the last supper: utility of the DNA barcode library for bloodmeal identification in ticks. *Mol. Ecol. Resour.* 12, 646–652
62 Hebert, P.D.N. *et al.* (2003) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proc. Biol. Sci.* 270 (Suppl. 1), S96–S99
63 Olson, Z.H. *et al.* (2012) An eDNA approach to detect eastern hellbenders (*Cryptobranchus a. alleganiensis*) using samples of water. *Wildlife Res.* 39, 629–636
64 Wilcox, T.M. *et al.* (2013) Robust detection of rare species using environmental DNA: the importance of primer specificity. *PLoS ONE* 8, e59520
65 Lodge, D.M. *et al.* (2012) Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Mol. Ecol.* 21, 2555–2558
66 Reichhardt, T. (1999) It's sink or swim as a tidal wave of data approaches. *Nature* 399, 517–520
67 Sun, Y. *et al.* (2012) A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis. *Brief. Bioinform.* 13, 107–121
68 Zhu, L. *et al.* (2011) Significant genetic boundaries and spatial dynamics of giant pandas occupying fragmented habitat across southwest China. *Mol. Ecol.* 20, 1122–1132
69 Schnell, I.B. *et al.* (2010) Characterisation of insect and plant origins using DNA extracted from small volumes of bee honey. *Arthropod Plant Interact.* 4, 107–116
70 Dalén, L. *et al.* (2007) Recovery of DNA from footprints in the snow. *Can. Field Nat.* 121, 321–324

71 Scandura, M. *et al.* (2006) An empirical approach for reliable microsatellite genotyping of wolf DNA from multiple noninvasive sources. *Conserv. Genet.* 7, 813–823
72 Willerslev, E. *et al.* (2007) Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* 317, 111–114
73 Epp, L.S. *et al.* (2012) New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Mol. Ecol.* 21, 1821–1833
74 Willerslev, E. *et al.* (2004) Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol. Evol.* 19, 141–147
75 Nichols, R.V. *et al.* (2012) Browsed twig environmental DNA: diagnostic PCR to identify ungulate species. *Mol. Ecol. Resour.* 12, 983–989
76 Folloni, S. *et al.* (2012) Detection of airborne genetically modified maize pollen by real-time PCR. *Mol. Ecol. Resour.* 12, 810–821
77 Walker, S.F. *et al.* (2007) Environmental detection of *Batrachochytrium dendrobatidis* in a temperate climate. *Dis. Aquat. Organ.* 77, 105–112
78 Knapp, M. *et al.* (2012) Generating barcoded libraries for multiplex high-throughput sequencing. *Methods Mol. Biol.* 840, 155–170
79 Knapp, M. and Hofreiter, M. (2010) Next generation sequencing of ancient DNA: requirements, strategies and perspectives. *Genes* 1, 227–243
80 Gansauge, M-T. and Meyer, M. (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* 8, 737–748
81 Mason, V.C. *et al.* (2011) Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Res.* 21, 1695–1704
82 Taberlet, P. *et al.* (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* 21, 2045–2050
83 Vestheim, H. and Jarman, S.N. (2008) Blocking primers to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. *Front. Zool.* 5, 12
84 Zhou, X. *et al.* (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *Gigascience* 2, 1–12
85 Ribeiro, F.J. *et al.* (2012) Finished bacterial genomes from shotgun sequence data. *Genome Res.* 22, 2270–2277
86 Schneider, G.F. and Dekker, C. (2012) DNA sequencing with nanopores. *Nat. Biotechnol.* 30, 326–328
87 Takahara, T. *et al.* (2012) Estimation of fish biomass using environmental DNA. *PLoS ONE* 7, e35868
88 Darling, J. and Mahon, A. (2011) From molecules to management: adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environ. Res.* 111, 978–988
89 Allentoft, M.E. *et al.* (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. Biol. Sci.* 279, 4724–4733
90 Ivanova, N.V. and Kuzmina, M.L. (2013) Protocols for dry DNA storage and shipment at room temperature. *Mol. Ecol. Resour.* 13, 890–898
91 Pietramellara, G. *et al.* (2009) Extracellular DNA in soil and sediment: fate and ecological relevance. *Biol. Fertil. Soils* 45, 219–235
92 Arnold, L.J. *et al.* (2011) Paper II – dirt, dates and DNA: OSL and radiocarbon chronologies of perennially frozen sediments in Siberia, and their implications for sedimentary ancient DNA studies. *Boreas* 40, 417–445
93 Biggs, J. *et al.* (2014) Analytical and methodological development for improved surveillance of the Great Crested Newt. Defra Project WC1067, Freshwater Habitats Trust, Oxford.

Appendix 2

Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter

Spens, J., Evans, A.R., Halfmaerten, D., Knudsen, S.W., Sengupta, M.E., Mak, S.S., Sigsgaard, E.E. and Hellström, M., 2017. Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods in Ecology and Evolution*, 8(5), pp.635-645v

Methods in Ecology and Evolution

Methods in Ecology and Evolution 2016 doi: 10.1111/2041-210X.12683

Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter

Johan Spens[1,2]†, Alice R. Evans[1]†, David Halfmaerten[3], Steen W. Knudsen[1], Mita E. Sengupta[4], Sarah S. T. Mak[1], Eva E. Sigsgaard[1] and Micaela Hellström[1,5]*

[1]*Centre for GeoGenetics, Natural History Museum of Denmark, Øster Voldgade 5-7, 1350 Copenhagen K, Denmark;* [2]*Wildlife, Fish and Environmental Studies, Swedish University of Agricultural Sciences, Skogsmarksgränd, 90183 Umeå, Sweden;* [3]*Research Institute for Nature and Forest, Gaverstraat 4, 9500 Geraardsbergen, Belgium;* [4]*Department of Veterinary Disease Biology Parasitology and Aquatic Diseases, Dyrlægevej 100, 1870 Frederiksberg C, Copenhagen, Denmark; and* [5]*Department of Ecology, Environment and Plant Sciences, Stockholm University, 10691 Stockholm, Sweden*

Summary

1. Aqueous environmental DNA (eDNA) is an emerging efficient non-invasive tool for species inventory studies. To maximize performance of downstream quantitative PCR (qPCR) and next-generation sequencing (NGS) applications, quality and quantity of the starting material is crucial, calling for optimized capture, storage and extraction techniques of eDNA. Previous comparative studies for eDNA capture/storage have tested precipitation and 'open' filters. However, practical 'enclosed' filters which reduce unnecessary handling have not been included. Here, we fill this gap by comparing a filter capsule (Sterivex-GP polyethersulfone, pore size 0·22 μm, hereafter called SX) with commonly used methods.

2. Our experimental set-up, covering altogether 41 treatments combining capture by precipitation or filtration with different preservation techniques and storage times, sampled one single lake (and a fish-free control pond). We selected documented capture methods that have successfully targeted a wide range of fauna. The eDNA was extracted using an optimized protocol modified from the DNeasy® Blood & Tissue kit (Qiagen). We measured total eDNA concentrations and Cq-values (cycles used for DNA quantification by qPCR) to target specific mtDNA cytochrome *b* (cyt *b*) sequences in two local keystone fish species.

3. SX yielded higher amounts of total eDNA along with lower Cq-values than polycarbonate track-etched filters (PCTE), glass fibre filters (GF) or ethanol precipitation (EP). SX also generated lower Cq-values than cellulose nitrate filters (CN) for one of the target species. DNA integrity of SX samples did not decrease significantly after 2 weeks of storage in contrast to GF and PCTE. Adding preservative before storage improved SX results.

4. In conclusion, we recommend SX filters (originally designed for filtering micro-organisms) as an efficient capture method for sampling macrobial eDNA. Ethanol or Longmire's buffer preservation of SX immediately after filtration is recommended. Preserved SX capsules may be stored at room temperature for at least 2 weeks without significant degradation. Reduced handling and less exposure to outside stress compared with other filters may contribute to better eDNA results. SX capsules are easily transported and enable eDNA sampling in remote and harsh field conditions as samples can be filtered/preserved on site.

Key-words: capsule, eDNA capture, environmental DNA, extraction, filter, monitoring, quantitative PCR, species-specific detection, water sampling method

Introduction

The realization that DNA from macrobiota can be obtained from environmental samples (environmental DNA, eDNA) started with excrements (Höss *et al.* 1992) and sediments (Willerslev *et al.* 2003). Over the last decade, the potential of aqueous eDNA to identify a wide range of plants and animals from a small volume of water has been realized (Martellini, Payment & Villemur 2005; Thomsen *et al.* 2012; Rees *et al.* 2014). Aqueous eDNA is an emerging increasingly sensitive technique for revealing species distributions (e.g. Jane *et al.* 2015; Valentini *et al.* 2016), early detection of invasive species (e.g. Smart *et al.* 2015; Simmons *et al.* 2016) and monitoring rare and/or threatened species for conservation (e.g. Zhan *et al.* 2013; McKee *et al.* 2015). Aqueous eDNA monitoring provides possibilities to upscale species distribution surveys considerably, because much less effort in time and resources are required compared to conventional methods (Dejean *et al.* 2012; Davy, Kidd & Wilson 2015). Based on literature

*Correspondence author. E-mail: micaela.hellstrom@su.se
†Joint first authors.

searches, we catalogue 49 studies successfully applying eDNA from water samples to detect macro-organisms in aquatic ecosystems, published between January 2005 and March 2015 (when this study was initiated; Table S1, Supporting Information). To our knowledge, 39 additional empirical studies were published since then, indicating a rapid rise of interest in this research area (Table S2).

The field of eDNA is still evolving, and a consensus of capture, storage and extraction methods has not yet been reached (Goldberg, Strickler & Pilliod 2015; Tables S1 and S2). In fact, the diversity of methods is almost as high as the number of research groups investigating this fairly new field of research. To ensure reliable results of downstream applications such as quantitative PCR (qPCR) and next-generation sequencing (NGS), the quantity and quality of the starting material is crucial. From our eDNA laboratory experience, we find that a modified easy-to-follow extraction protocol resulting in high yields is needed. Based on eDNA studies published so far (Tables S1 and S2), we identify three pre-PCR key issues that hold opportunities for improvement: (i) capturing sufficient quantities of eDNA as quite a few studies report low amounts of captured total eDNA, (ii) effectively preserving eDNA samples before extraction and (iii) lowering contamination risks from collection to extraction of eDNA.

Comparative studies on aqueous eDNA capture and storage techniques (i.e. optimal ways of preserving the eDNA captured on the filters until extraction; e.g. Renshaw *et al.* 2015) were based on the so-called 'open filters' (requiring handling, a filter funnel and a vacuum pump; e.g. Liang & Keeley 2013; Turner *et al.* 2014b) and ethanol precipitation (EP; e.g. Piaggio *et al.* 2014; Deiner *et al.* 2015). However, no enclosed filters were included in previous comparative assays.

The Sterivex-GP capsule filter (SX), with a polyethersulfone membrane, is a standard method for characterizing microbial communities (Chestnut *et al.* 2014) and for removing pathogens from water as the organisms are captured on the filter membranes. To our knowledge, only two published aqueous eDNA studies have used this filter to detect aquatic macro-organisms (fish detection: Keskin 2014; Bergman *et al.* 2016), and the technique has been successful to detect a wide range of aquatic macro-organisms in Denmark and Belgium (M. Hellström, M.E. Sengupta, S.W. Knudsen, D. Halfmarten. unpublished, S1). The SX filter is enclosed in a capsule, which reduces handling. A water sample can easily be filtered in the field, saving time and facilitating fixation of the eDNA immediately after capture. Additionally, downstream DNA extraction takes place within the filter capsules with no need for the membrane to be removed or handled. We therefore test the performance of SX compared to other more frequently used eDNA capture methods (Table S1), under different storage conditions, in an effort to address issues 1–3 above. To date, there are no studies comparing SX to other capture methods and multiple storage treatments. We aim to fill this gap, with an experimental study comparing SX with four other capture methods in a set-up with five typical storage treatments and three different storage times (up to 2 weeks). The tested open filter materials polycarbonate, cellulose nitrate and glass fibre (GF) and the range of tested pore sizes (0·2–0·6 μm) are typical of previous studies (Tables S1 and S2). We used an optimized extraction protocol based on a commercial kit to increase eDNA yields. To evaluate the usefulness of the SX and preservation buffers in comparison with typically used methods (Tables S1 and S2), we test the following H_0 hypotheses:

$H_0$1. CAPTURE METHOD: SX is equally effective as other tested eDNA capturing techniques in regard to DNA quantity and quality measured as the total extracted eDNA concentration [$eDNA_{tot}$] and as Cq-values (quantification cycles, *sensu* Bustin *et al.* 2009) from two species-specific qPCR assays.

$H_0$2a. STORAGE PRESERVATIVE: Storing filters with a preservation buffer does not affect qPCR amplification compared to immediate extraction or freezing at −20 °C (no buffer added).

$H_0$2b. STORAGE TIME: There is no significant difference in eDNA quality over time between SX and the other tested capturing techniques.

$H_0$3. CONTAMINATION: There is no significant difference between SX and the other tested capture techniques in occurence of false positives.

To test these hypotheses, we use an experimental set-up with subsampling a single large homogenous sample of water from a Danish lake. Subsamples are subjected to different eDNA capture methods within the same day followed by different storage treatments. A control site (fish-free pond) is sampled using the same set-up. Each capture and storage treatment is assessed using concentration of total eDNA as well as species-specific qPCR assays targeting pike *Esox lucius* L. and perch *Perca fluviatilis* L. By testing H_0 hypotheses (1–3), the multiple opportunities for optimization of eDNA surveys held by the use of SX may be empirically evaluated. Based on the results, we suggest recommendations for improved capture, storage and extraction to use for aqueous eDNA, taking remote and harsh field conditions into consideration.

Materials and methods

STUDY SITES

We chose Gentofte Lake, Denmark (N55·7435°, E12·5348°), as the study site and a fish-free pond in Copenhagen botanical garden as a negative field control (N55·6875°, E12·5746°). Gentofte Lake (26 ha) is an alkaline clear water (Appendix S2) harbouring a wide range of fish species, including pike and perch.

WATER COLLECTION

We retrieved 130 L of water from Gentofte Lake on 17 March 2015. The water (4 °C) was collected at *c.* 30 points along *c.* 100 m of shoreline close to the outlet of the lake. Additionally, we collected 40 L of water from the control pond on 21 March 2015. The water was

collected in sterilized 5-L buckets which prior to sampling were soaked in bleach (5%) for 10 min, and then rinsed with laboratory-grade ethanol (70%). The containers were soaked repeatedly in lake water at a location away from the collection point. Nitrile gloves were used during cleaning, collection and filtration.

CAPTURE AND STORAGE

We carried out 41 different treatment combinations of the water sample in total (Table 1, Fig. S1). We used five capture techniques, five storage methods and three time regimes. All treatments were performed in triplicate. Apart from an in-house modified SX procedure (see Fig. 1), the capture and storage methods were based on published sources (Table S1). The capture methods (hereafter referred to with their abbreviations in square brackets) were as follows: (i) ethanol precipitation [EP] (Ficetola *et al.* 2008), (ii) mixed cellulose esters membrane filters including cellulose nitrate and cellulose acetate [CN]; Advantec 47 mm diameter 0·45 μm pore size (Toyo Roshi Kaisha, Ltd., Tokyo, Japan), (iii) polycarbonate track-etched filters [PCTE]; Whatman Nucleopore Membrane 47 mm diameter 0·2 μm pore size (Merck KGaA, Darmstadt, Germany)], (iv) glass fibre [GF] membrane filters; Advantec GA-55 47 mm diameter 0·6 μm pore size (Toyo Roshi Kaisha, Ltd., Tokyo, Japan) and (v) sterivex-GP capsule filters [SX]; polyethersulfone 0·22 μm pore size with luer-lock outlet (Merck KGaA)]. Further downstream, SX was divided into an extraction from the filter within the capsule ($SX_{CAPSULE}$), after removal of the storage buffer, and an extraction from the removed preservation buffer within a centrifuge tube (SX_{TUBE}; see DNA extraction section below). The different storage methods were as follows: (i) ethanol 99% 200 proof at room temperature (RT), Molecular Biology Grade (Thermo Fisher Scientific Inc., Waltham, MA, USA); (ii) Longmire's buffer at RT (Longmire's; Longmire, Maltbie & Baker 1997); (iii) RNAlater at RT (RNA Stabilization Reagent; QIAGEN, Stockach, Germany); (iv) no buffer, frozen at −20 °C; and (v) no buffer, refrigerated at 8–10 °C. The three time regimes between filtration and extractions were (i) within 5 hours (5 h), (ii) within 24 h and (iii) after 2 weeks. Each treatment (n = 41) was performed in triplicate. For each filter replicate, 1 L of lake water was processed (0·015 L for EP). For each capture–storage treatment, we included one negative control without lake water. Additionally, 1 L tap water was run through each filter (0·015 L for EP) as a control to detect potential contamination from the filtration facilities. For the control pond, one sample per capture–storage treatment was processed (n = 23). We captured eDNA from 155 subsamples and negative controls altogether. The water samples were filtered or ethanol-precipitated by a team of 10 researchers and the replicates of each treatment started at different times to avoid temporal bias of filtrations. Prior to DNA capture, bench surfaces and all equipment were wiped with bleach (5%) and laboratory-grade ethanol (70%). Prior to each collection of subsamples, the water was mixed thoroughly in the 130-L container. For the open membrane filter (GF, CN and PCTE), 1 L water samples were vacuum-filtered (*c.* 15–30 min) using Nalgene 250-mL sterile disposable test filter funnels (Thermo Fisher Scientific Inc. USA). The filters were removed from the funnel with forceps and then placed in 5-mL DNA LoBind® centrifuge tubes (Eppendorf AG, Hamburg, Germany) that were either empty (if the time regime was 5 h or the storage method was freezing) or contained preservation buffer. For all treatments and downstream applications, Eppendorf DNA LoBind® tubes were used in order to avoid up to 50% retention of DNA by the plastic, which is a documented problem especially for short DNA fragments (Gaillard & Strauss 1998; Ellison *et al.* 2006). For the SX filters, 1 L of water was slowly (*c.* 10 min to avoid tearing of filters, following manufacturer's recommendations) pushed through each filter capsule using a prepacked sterile 50-mL luer-lock syringe. Remaining water in the SX was removed by pushing air through the filter until dry, also using the syringe. The outlet ends of the filters were closed with MoBio outlet caps (MOBIO Laboratories, QIAGEN) and 2 mL preservation buffer was pipetted to the inlet end using filter tips. The inlet ends were closed with inlet caps (MOBIO Laboratories, QIAGEN) and both ends were sealed with parafilm whereafter the capsules were inverted vigorously. The frozen samples and the (5 h) and (24 h) EP samples were placed at −20 °C until extraction, while the non-treated samples (5 h) were placed in a refrigerator and extracted directly after the filtering session. Samples containing buffers were stored at RT until processed. The (2 weeks) EP samples were frozen for 24 h prior to extraction to allow for precipitation. In total, we processed 96·135 L of water from the lake (32 treatments × 3 replicates × 1 L + 3 EP treatments × 3 replicates × 0·015 L) and 20·045 L of water from the control pond (20 treatments × 1 replicate × 1 L + 3 EP treatments × 1 replicates × 0·015 L; Table 1).

MOLECULAR LABORATORY CONDITIONS

DNA extractions and qPCR assays took place in the laboratories at the Centre for GeoGenetics, University of Copenhagen, Denmark. The facilities are designed for handling environmental samples requiring the most stringent precautions to avoid contamination. Pre-PCR, extraction and PCR facilities are located in separate designated rooms with positive air pressure. Laboratory coats are changed between rooms. Prior to any work in the laboratory, all surfaces are washed with 5% bleach and 70% ethanol. After completing extractions

Table 1. Outline of the number of samples processed per capture and storage treatment (negative control pond in parentheses)

		Storage								
			Frozen	Ethanol	Longmire's	RNAlater	Frozen	Ethanol	Longmire's	RNAlater
Capture	Sum	Refrigerated 5 h	24 h				2 weeks			
$SX_{CAPSULE}$	27 (5)	3 (1)	3 (1)	3 (1)	3 (1)	3 (1)	3	3	3	3
SX_{TUBE}	18 (3)			3 (1)	3 (1)	3 (1)		3	3	3
Cellulose nitrate	15 (5)	3 (1)	(1)	(1)	(1)	(1)	3	3	3	3
Glass fibre	27 (5)	3 (1)	3 (1)	3 (1)	3 (1)	3 (1)	3	3	3	3
Polycarbonate	27 (5)	3 (1)	3 (1)	3 (1)	3 (1)	3 (1)	3	3	3	3
Precipitation	9 (3)	3		3 (3)				3		
Total	123 (26)									

Sterivex, eDNA extraction within capsule ($SX_{CAPSULE}$); Sterivex, eDNA extraction from buffer in tube outside capsule (SX_{TUBE}).

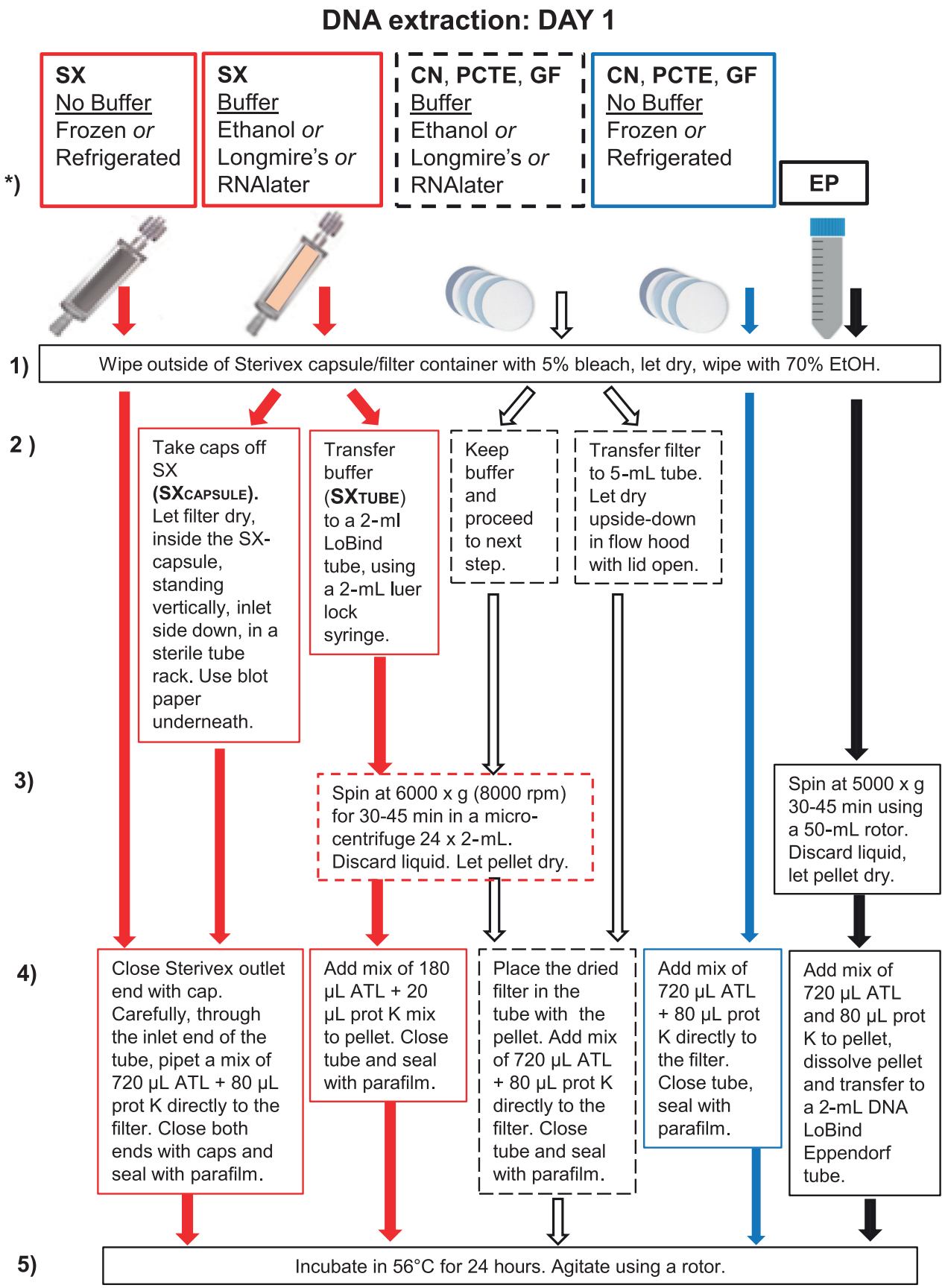


Fig. 1.

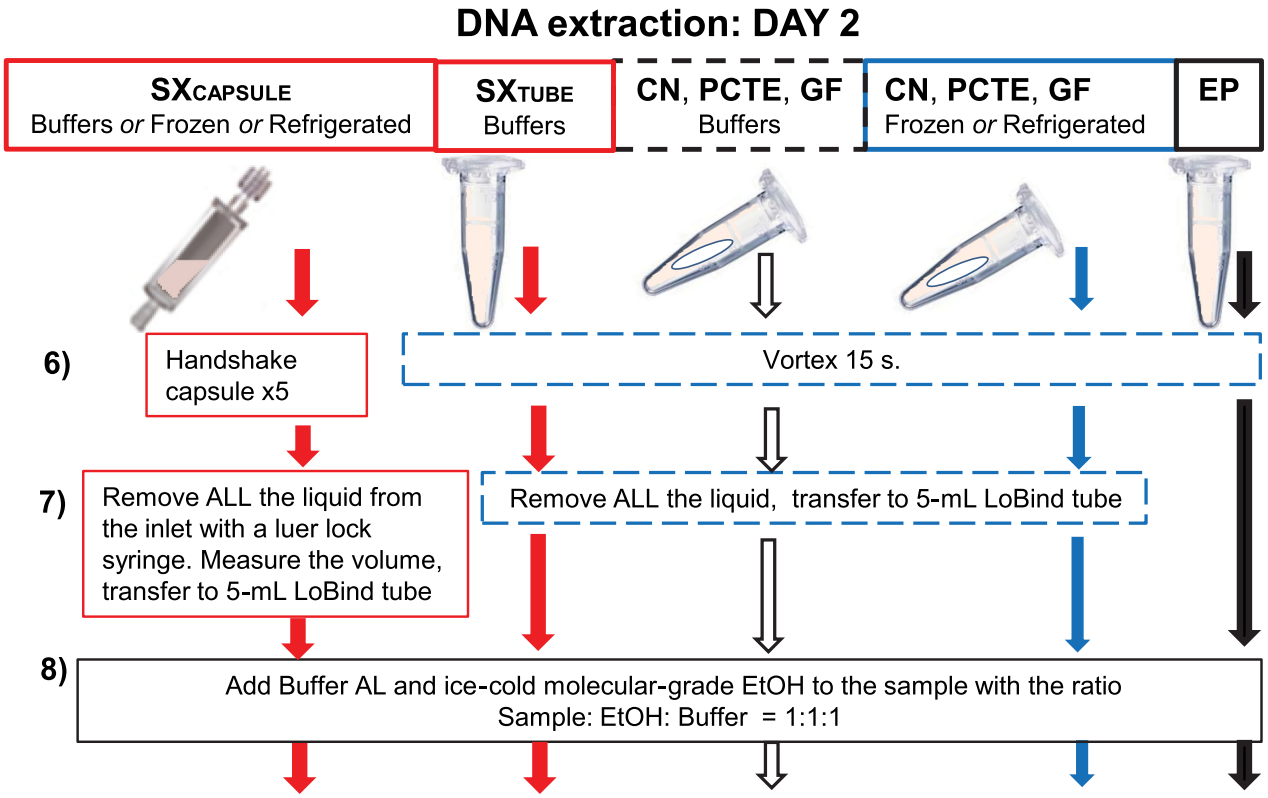


Fig. 1. Flow chart illustrating the modified environmental DNA (eDNA) extraction protocol based on DNeasy Blood & Tissue Kit (QIAGEN, Carlsbad, CA, USA). *) Capture: SX, Sterivex-GP polyethersulfone capsule filters, Note that $SX_{CAPSULE}$ and SX_{TUBE} are treated as separate samples from step 2. CN, cellulose nitrate; PCTE, polycarbonate track-etched; GF, glass fibre filters; EP, ethanol precipitation. Storage: Frozen at −20 °C, Refrigerated are samples stored at 8–10 °C and processed within 5 h. Steps 9–26 see Appendix S1.

involving guanidiumthiocyanate, surfaces are washed with 70% ethanol (to avoid reactions between chlorine in the bleach and guanidiumthiocyanate in two of the buffers provided with the Qiagen kit), 5% bleach and then 70% ethanol. All extractions of eDNA took place in laminar flow hoods which were UV-treated before and after extractions. Every night, the entire facilities are automatically UV-treated for a 2-h period.

DNA EXTRACTION

We extracted the eDNA using the extraction protocol outlined in Fig. 1 and Appendix S1. The SX filters containing preservation buffers underwent two extractions, one extraction from the buffer and one extraction within the filter capsule after it had been emptied of buffer (hereafter referred to as SX_{TUBE} and $SX_{CAPSULE}$). Altogether, 179 (24 SX_{TUBE} + 155 (see 'Capture and storage' section above) samples from the study lake and the control pond were extracted. We measured [$eDNA_{tot}$] in each extraction using a Qubit 1.0 fluorometer (Thermo Fisher Scientific Inc.) applying the high-sensitivity assay for dsDNA (Life Technologies, Carlsbad, CA, USA).

QUANTITATIVE PCR

For the qPCR assays (e.g. Wilcox *et al.* 2013), two species-specific TaqMan primers/probe sets were used targeting 84 and 89 base pair fragments of the mitochondrial cytochrome *b* (cyt *b*) gene in pike and perch, respectively (Table S3). Species specificity of the assays was tested on extracted DNA from non-target species (Table S3) using the qPCR set-up described below. These non-target species did not generate any amplification signals. The optimal ratio of probe: primer concentration was tested prior to the study. The final PCR set-up to detect the target species was as follows: pike – 5 μL template DNA, 12·5 μL TaqMan Environmental Master Mix 2·0 (Life Technologies), 3 μL forward primer (10 μM), 2 μL reverse primer (10 μM) and 3 μL probe (2·5 μM); and perch – 5 μL template DNA, 12·5 μL TaqMan Environmental Master Mix 2·0 (Life Technologies), 0·5 μL forward primer (10 μM), 2·5 μL reverse primer (10 μM), 3 μL probe (2·5 μM) and 1·5 μL UV-treated laboratory-grade water. The TaqMan qPCRs were performed on a Stratagene Mx3005P (Thermo Fisher Scientific Inc.) using thermal cycling parameters of 50 °C (5 min), 95 °C (10 min) followed by 50 cycles of 95 °C (30 s) and 60 °C (1 min). For each plate, no-template controls (NTCs) and positive/negative tissue extracts were run alongside the samples. All filtering and extraction negatives were included in the qPCR assays. Additional qPCR replicates were run in order to detect effects of freezing and thawing of the samples. To check for PCR inhibition in the lake, separate qPCR assays for both species following the protocols above were performed in a dilution series (1 : 1, 1 : 2, 1 : 10 and 1 : 20) of extracted DNA on four samples replicated twice plus two positive and two negative controls to determine any deviation of the amplification curves. The dilution series did not indicate inhibition.

DATA ANALYSIS

To compare detection probability (i.e. diagnostic sensitivity) between eDNA capture methods, the proportion of positive qPCR replicates was calculated for each target species. Positive samples were analysed

using multivariate decision trees and univariate tests of 'no-effect' null hypotheses. To explore the effect of capture and storage on qPCR Cq-values, Chi-square Automatic Interaction Detector (CHAID) decision tree was used. CHAID is a nonparametric tree-building method that can handle multivariate categorically induced quantitative responses (IBM Corp. (2013)). It defines optimal multiway splits and adjusts for Bonferroni. The main advantage of this approach is to analyse a data set all-in-one (rather than manually splitting the data into user-selected subgroups and thereafter choosing and performing multiple tests). The approach offers a number of other advantages including its ability to handle categorical (ordered, nominal) data types well and to model nonlinear relationships without having to specify *a priori* the form of the interactions. A CHAID tree produces an overview, grouping or singling out the factors that predict the variation in the response variable. Categorical variables (capture method, storage treatment and storage time) were used as model predictors, and Cq-value from qPCR was set as the response target. Two trees were generated: the first targeting perch and the second pike. Tree depth, that is the maximum number of branching levels, was set to two (realized from ten 50/50 split validations) to reduce overfitting.

For a univariate test of H_0 (1–2a,b), first a Wilcoxon signed-rank test for paired samples was applied to determine whether [$eDNA_{tot}$] and Cq-values attained using $SX_{CAPSULE}$ differ significantly, from any of the other tested capture methods (CN, GF, PCTE, EP and SX_{TUBE}). Secondly, SX, GF and PCTE filter results were tested for signs of eDNA degradation over time, that is detecting any significant difference in Cq-values or [$eDNA_{tot}$] between 24 h and 2 weeks of storage. Wilcoxon signed-rank test was used as data exhibited non-normal distributions. Thirdly, guided by results from the CHAID trees, results from $SX_{CAPSULE}$ stored in ethanol or Longmire's were tested (Mann–Whitney) for differences in Cq-value against $SX_{CAPSULE}$ without preservation buffer. The CN filter group was reduced, as the planned 1-day storage treatment was omitted due to filtering time constraints. The mean difference in Cq-value and associated 95% CI of all qPCR replicates was calculated. All statistical analyses were performed using SPSS IBM Corp. (2013).

Results

SPECIES DETECTION

Altogether 713 qPCR samples, including controls, were analysed. No samples were discarded. Perch and pike were both detected in most of the qPCR runs from the study lake (314 of 365, Fig. 2). For both species, SX_{TUBE} showed the highest overall detection rate (95% perch and 96% pike) and EP the lowest (89% perch and 56% pike; overall difference $SX_{TUBE} \neq$ EP: Pearson χ^2 (1, n = 62) = 6·9, Fisher's exact P = 0·02).

CAPTURE METHOD

A CHAID tree multivariate predictive model was successfully generated from perch Cq-values. Capture method was the best overall predictor of Cq-values, better than storage media or storage time. In general, the lowest Cq-values were generated from $SX_{CAPSULE}$ samples in comparison with other capture methods (Fig. 3a). We validated the fundamental first-level outcome from this multivariate model for perch with new data in the build of a second CHAID tree, modelling pike Cq-values (Fig. 3b). In this second variant, capture was also the best predictor of Cq-values and $SX_{CAPSULE}$ tied with the CN and GF filters in the lowest value category.

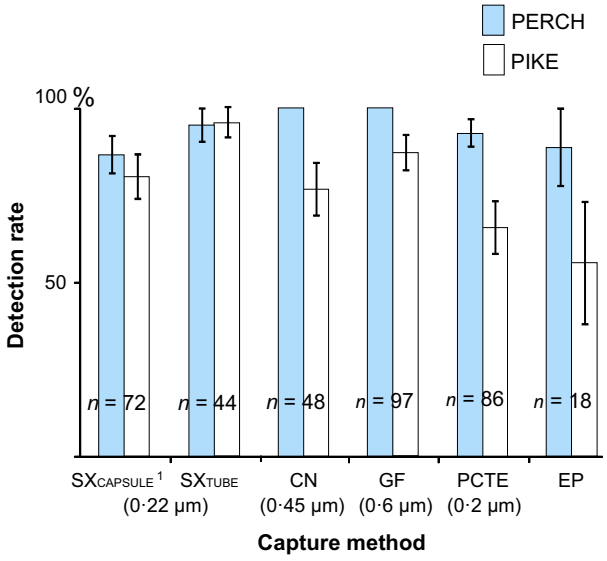


Fig. 2. Detection rate using quantitative PCR (qPCR; study lake). Blue bars and clear bars show positive detections of perch and pike, respectively. Pore size of filters within parentheses. $SX_{CAPSULE}$, Sterivex, extraction within filter capsule; SX_{TUBE}, Sterivex, extraction in tube outside capsule from removed preservation buffer; CN, cellulose nitrate; PCTE, polycarbonate track-etched; GF, glass fibre; EP, ethanol precipitation. Error bars represent standard errors; n indicates number of trials pooling all replicates for each method and both species combined. [1]Deviating from protocol, 12 $SX_{CAPSULE}$ replicates were over-vortexed and tested mainly negative. If these 12 over-vortexed samples are omitted, the detection rate estimate for $SX_{CAPSULE}$ increases to 100% for perch and to 91% for pike.

The fundamental first-level outcome of both the CHAID tree multivariate predictive models was supported in a one-by-one comparison of capture methods including both species and all treatments. Overall, $SX_{CAPSULE}$ was more efficient than the other capture methods apart from CN. $SX_{CAPSULE}$ yielded significantly higher [$eDNA_{tot}$] and lower Cq-values (Table 2). SX samples contained up to 118 ng total eDNA μL^{-1} and most $SX_{CAPSULE}$ amplified before 36 cycles (Fig. 4). [$eDNA_{tot}$] from the fish-free control pond showed a similar pattern, being higher for CN and $SX_{CAPSULE}$ compared with GF and PCTE (Mann–Whitney U = 12, $n_1 = n_2$ = 10, Fisher's exact P = 0·003), but with no Cq-values from qPCR as target species were not present. Overall, capture method and [$eDNA_{tot}$] were fundamental predictors of Cq-values (Fig. 4).

STORAGE PRESERVATIVE

SX-specific storage results are singled out and illustrated in Fig. 5. SX_{TUBE} samples treated with RNAlater, a significant predictor of poorer Cq-values in the CHAID trees, were least successful. For $SX_{CAPSULE}$, preservation in ethanol or Longmire buffer improved Cq-values for perch in comparison with frozen, 5 h and preservation in RNAlater (Figs 3a and 6). Also for both species pooled, these two buffers (ethanol or

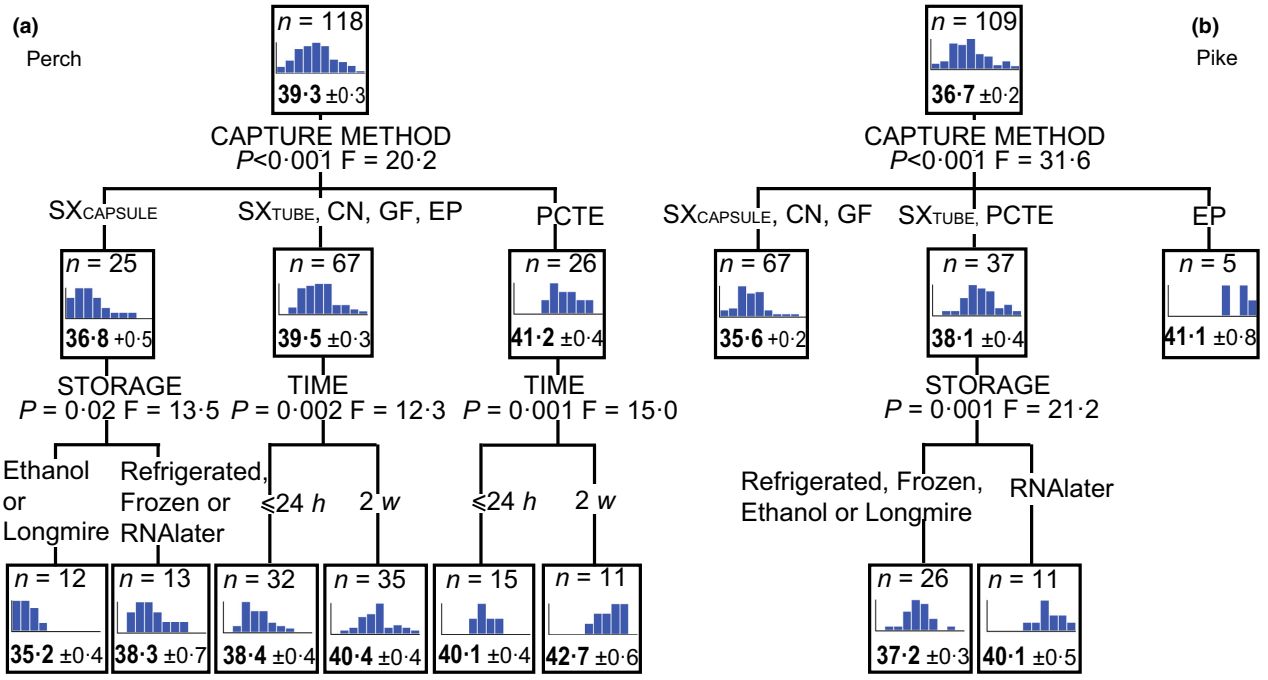


Fig. 3. Chi-square Automatic Interaction Detector decision trees relating three categorical variables (capture method, storage treatment and storage time) as model predictors for Cq-values as response target (study lake). (a) Perch. Best predictor was capture method, followed by storage time, and finally, storage treatment. (b) Pike. Best predictor was capture method followed by storage treatment. $SX_{CAPSULE}$, Sterivex, extracted within capsule; SX_{TUBE}, Sterivex, extraction in tube outside capsule; CN, cellulose nitrate; GF, glass fibre; PCTE, polycarbonate track-etched fibre; EP, ethanol precipitation; h, hours; w, weeks. Blue bar charts indicate relative size distribution of Cq-values within each category before split. Number under bar charts indicate mean Cq-value for the given category ± SE.

Table 2. $SX_{CAPSULE}$ in comparison with other eDNA capture methods

$SX_{CAPSULE}$ comparison of Cq-values ($SX_{CAPSULE}$ comparison of [$eDNA_{tot}$])					
Capture	Pairs of *n*	*P*	Significance*	*Z*	Rank
SX_{TUBE}	33 (18)	**1×10^{-5} (5×10^{-4})**	*** (**)	−4·4 (−3·5)	$SX_{CAPSULE}$ < SX_{TUBE} (>SX_{TUBE})
GF	50 (27)	**7×10^{-3} (2×10^{-5})**	* (***)	−2·7 (−4·3)	$SX_{CAPSULE}$ < GF (>GF)
PCTE	44 (27)	**1×10^{-5} (6×10^{-6})**	*** (***)	−4·4 (−4·5)	$SX_{CAPSULE}$ < PCTE (>PCTE)
EP	13 (9)	**1×10^{-3} (8×10^{-3})**	** (*)	−3·2 (−2·7)	$SX_{CAPSULE}$ < EP (>EP)
CN†	29 (15)	0·32 (0·55)	N.S. (N.S.)	−1·0 (−0·6)	

Wilcoxon matched-pair signed-rank test of both Cq-values from qPCR and [$eDNA_{tot}$] (denoted in parentheses). Significant *P*-values are in bold and non-significant *P*-values are denoted as N.S.

$SX_{CAPSULE}$, Sterivex, extracted within capsule; SX_{TUBE}, Sterivex, extraction in tube outside capsule; GF, glass fibre; PCTE, polycarbonate track-etched filter; CN, cellulose nitrate; EP, ethanol precipitation; [$eDNA_{tot}$], total eDNA concentration.

*Bonferroni corrected (5 tests): $\alpha = 0{\cdot}05$ lowered to 0·01, $\alpha = 0{\cdot}01$ lowered to 0·002 and $\alpha = 0{\cdot}001$ lowered to 0·0002.

†Due to time constraints, CN (24 h) were cancelled reducing sample size and statistical power for CN in comparison.

Longmire) in $SX_{CAPSULE}$ resulted in lower Cq-values compared with frozen or 5 h (Mann–Whitney Test U: 35, $n_1 = 23$, $n_2 = 15$, $Z = -4{\cdot}1$; $P = 4 \times 10^{-5}$).

STORAGE TIME

Storage time in the second-level outcome from the first CHAID tree was classified as a positively correlated predictor of Cq-values for all capture methods apart from SX (Fig. 3a). This was supported in a one-by-one comparison of capture methods including both species and 24 h to 2 weeks treatments (Table 3). Cq-values did not increase significantly with time using SX, but did with GF and PCTE.

The mean difference between Cq-values of paired qPCR replicates run within the same day was +0·3 ± 0·2 SE. This difference increased to +1·3 ± 0·2 SE when replicates run on different days were included, indicating that freezing and thawing of eDNA once or twice between measurements decreased DNA quality [Welch's test $t(1, 68) = 7{\cdot}1$, $n_1 = 20$, $n_2 = 80$, $P = 9 \times 10^{-10}$]. To avoid introducing this error, only DNA templates thawed for the first time were included when calculating average Cq-values for the samples.

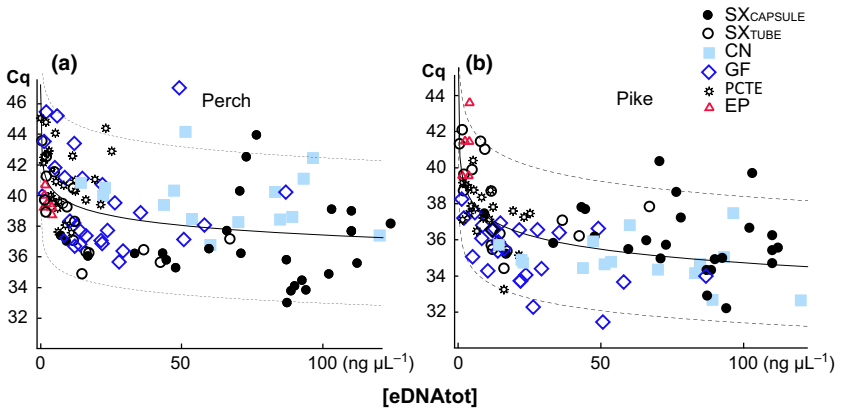


Fig. 4. Environmental DNA (eDNA) capture methods: relationship between total eDNA concentration ([$eDNA_{tot}$]) and quantification cycles in qPCR (Cq-value) in study lake. Line represents best-fit power function where Cq decreased as a function of [$eDNA_{tot}$]. (a) Perch. $Cq = 41{\cdot}8 \times [eDNA_{tot}]^{-0{\cdot}024}$; $P < 0{\cdot}001$, $R^2 = 0{\cdot}23$. (b) Pike: $Cq = 40{\cdot}0 \times [eDNA_{tot}]^{-0{\cdot}031}$; $P < 0{\cdot}001$, $R^2 = 0{\cdot}42$. Dotted lines represent lower or upper limits of 95% CI for slope of regression. $SX_{CAPSULE}$, Sterivex, extracted within capsule; SX_{TUBE}, Sterivex, extracted from buffer in tube outside capsule; CN, cellulose nitrate; GF, glass fibre; PCTE, polycarbonate track-etched fibre; EP, ethanol precipitation.

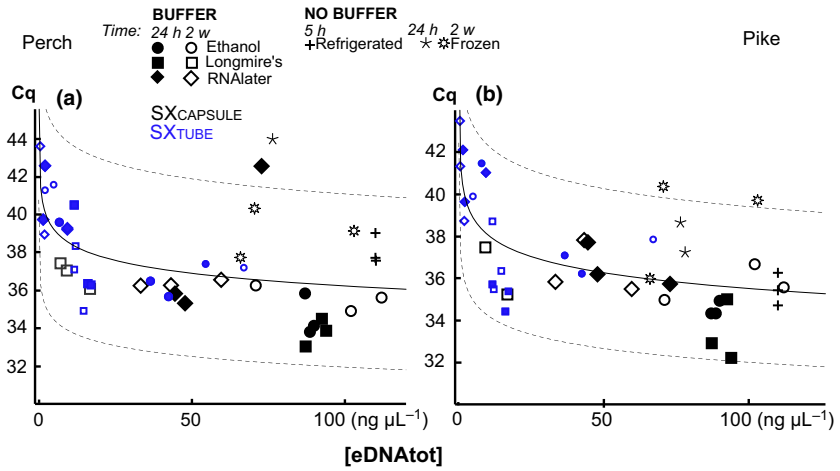


Fig. 5. Environmental DNA (eDNA) storage treatment using SX: relationship between total eDNA concentration ([$eDNA_{tot}$]) and quantification cycles in qPCR (Cq-value) in study lake. Line represents best-fit power function of the negative correlation between Cq and [$eDNA_{tot}$]. (a) Perch: $Cq = 40{\cdot}9 \times [eDNA_{tot}]^{-0{\cdot}026}$; $P < 0{\cdot}001$, $R^2 = 0{\cdot}28$. (b) Pike: $Cq = 40{\cdot}8 \times [eDNA_{tot}]^{-0{\cdot}030}$; $P < 0{\cdot}001$, $R^2 = 0{\cdot}45$. Dotted lines represent lower or upper limits of 95% CI for slope of regression. Sterivex, extracted within capsule ($SX_{CAPSULE}$) and from buffer in tube outside capsule (SX_{TUBE}) shown in black and blue symbols, respectively. h, hours; w, weeks.

CONTAMINATION

One false-positive signal for perch was detected at 42 cycles in an EP 'no-water' negative control. Remaining negative controls for capture/storage treatments ($n = 80$) and negative pond water ($n = 85$), NTCs ($n = 64$) and 37/40 tissue negative controls for species specificity did not amplify. The contaminated tissue control was replaced and showed no amplification. One extraction blank came up positive in one of the seven runs, but at a very high Cq of 46·2.

Discussion

To our knowledge, this is the first study comparing enclosed filters (SX) with commonly used eDNA capture and storage techniques. Similarly to other capture methods, SX can be used to target a wide range of macro-organisms successfully (using PCR, qPCR or NGS; Table S1), ensuring the generality of SX for surveys of aquatic biodiversity.

Specifically, SX with added preservation buffer (ethanol or Longmire's) is the optimal approach of the tested treatments in regard to [$eDNA_{tot}$] yield and detection sensitivity for target species. Other eDNA studies of macrobiota using SX (Keskin 2014; Bergman *et al.* 2016) did not apply preservation buffers. Although our study set-up was different, the lake sample results are consistent with the mesocosm experiment of Renshaw *et al.* (2015), showing that open CN filter and polyethersulfone filters (same material as SX in this study) were more effective than PCTE and GF. Additionally, we demonstrate that SX eDNA retains integrity over time, whereas eDNA from the open filters degrades significantly. These results suggest that SX eDNA is more effectively preserved, possibly due to the fact that it is considerably less handled by the user. The capsule may reduce risks of exposure to physical and biogenic stress as well as contamination, because capture, storage and extraction take place within the filter capsule. This, together with extended field usage possibilities, and higher eDNA yields, constitutes reasons to recommend enclosed filters before other capture methods.

CAPTURE METHOD

Based on our results, we reject H_0 hypothesis 1 stating that SX and commonly used techniques in our study are equally

effective, because $SX_{CAPSULE}$ yields the lowest Cq-values for perch (Fig. 3a). However, this is only partially validated in the case of pike (Fig. 3b), where $SX_{CAPSULE}$, GF and CN group together for the lowest Cq-values. Overall, $SX_{CAPSULE}$ yields higher [$eDNA_{tot}$] and generates better qPCR results than other capture methods, with the exception of CN. Our CN/SX comparisons are not as extensive as the SX/GF and SX/PCTE comparisons (Table 2). We show that higher levels of [$eDNA_{tot}$] are related to lower Cq-values of target species DNA (R^2 = 0·23–0·45, Figs 4 and 5) and therefore suggest measurements of [$eDNA_{tot}$] for approximate indications of eDNA capture efficiency.

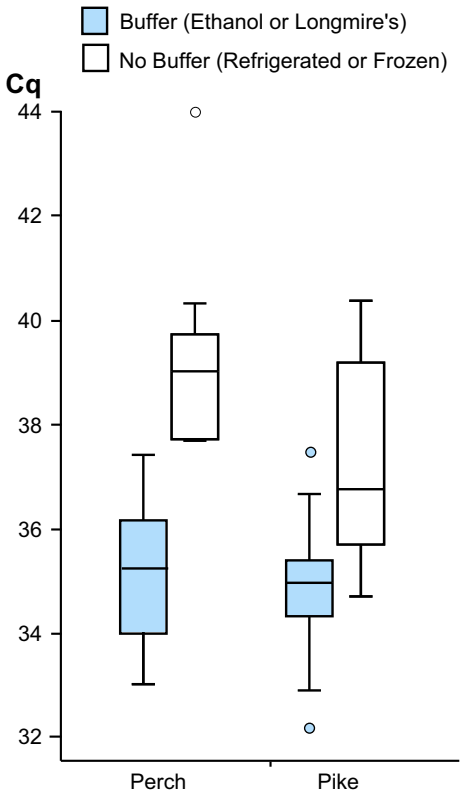


Fig. 6. Boxplots of Cq-values showing $SX_{CAPSULE}$ (extraction within Sterivex capsule) filter storage with and without preservation buffer (ethanol or Longmire's).

The comparison in this study of SX_{TUBE} to $SX_{CAPSULE}$ demonstrates that utilizing both these sources of eDNA should be useful. Pooling of these in the final elution step would be advisable for gaining even higher final yields of eDNA. SX_{TUBE} exhibits the highest overall detection rate for both species (95–96%) in our study, significantly higher than EP results. Higher amounts of false negatives from EP field samples may be due to DNA retention in the falcon tubes (Gaillard & Strauss 1998) and/or to the low water volume processed (0·015 L; Deiner *et al.* 2015; Eichmiller, Miller & Sorensen 2016; Minamoto *et al.* 2016).

STORAGE PRESERVATIVE

We reject H_0 hypothesis 2a stating that preservation buffers for storage of SX do not affect qPCR amplification in comparison with extraction within 5 h or freezing at −20 °C. Two-thirds of published aqueous eDNA surveys reporting storage details apply freezing of filters as a preservation method (Table S1 and S2), while less than one-third of surveys use buffer storage. Our results indicate that addition of ethanol or Longmire's immediately after SX filtration provides the lowest Cq-values, and is significantly better than freeze storage or extraction within 5 h. Based on our results as well as the results of three previous studies (Renshaw *et al.* 2015; Wegleitner *et al.* 2015; Minamoto *et al.* 2016), we recommend addition of preservation immediately after filtration.

STORAGE TIME

We reject H_0 hypothesis 2b that degradation of captured eDNA is the same in SX filters and the other capture techniques tested in this study. Cq-values increase significantly with storage time for GF and PCTE samples, indicating degradation of eDNA. In contrast, Cq-values for SX samples ($SX_{CAPSULE}$ or SX_{TUBE}) do not differ significantly after 2 weeks of storage at RT.

We note that repeated use of the same extracted eDNA sample (eluted in TE-buffer) for qPCR on different days, entailing repeated freezing and thawing, resulted in higher Cq-values. Freeze–thaw-induced degradation and/or inhibition of DNA is previously acknowledged (e.g. Ross, Haites

Table 3. Effect of storage time for eDNA results with different capture methods

Paired test of Cq-values					
Storage	Pairs of *n*	*P*	Significance*	*Z*	Rank
$SX_{CAPSULE}$	20	0·15	N.S.	−1·5	
SX_{TUBE}	16	0·18	N.S.	−1·3	
PCTE	16	**0·002**	**	−3·1	PCTE 24 h < PCTE 2 weeks
Glass fibre (GF)	24	**0·002**	**	−3·1	GF 24 h < GF 2 weeks

Wilcoxon matched-pair signed-rank test of Cq-values from qPCR. Storage 24 h paired with storage 2 weeks. Significant *P*-values are in bold and non-significant *P*-values are denoted as N.S.
Due to time constraints, cellulose nitrate treatments (24 h) were cancelled.
$SX_{CAPSULE}$, Sterivex, extracted within capsule; SX_{TUBE}, Sterivex, extraction in tube outside capsule; PCTE, polycarbonate track-etched filter.
*Bonferroni corrected (4 tests): α = 0·05 lowered to 0·0125, α = 0·01 lowered to 0·0025.

& Kelly 1990; Takahara, Minamoto & Doi 2015). We therefore recommend that extracted eDNA samples are divided into many aliquots immediately after extraction, in order to avoid compromising eDNA quality by repeated freezing and thawing.

CONTAMINATION

We cannot yet reject H_0 hypothesis 3 stating that SX leads to as many false positives as typically used methods. We only produced one false positive (EP) which is insufficient for any statistical inference. The SX approach using sealed pre-sterilized equipment until sampling, and capping filter immediately after filtration, should reduce contamination risk. The contamination variance between these capture methods remains to be tested using more observations and possibly synthetic controls (Wilson, Wozney & Smith 2016).

LIMITATIONS

The hand-held syringe used with SX filter units is convenient but turns into a labour-intensive bottleneck when processing many samples. This can be alleviated by switching to battery-powered pumps (Sterivex™ 2013). In 'algal soup' or turbid waters, 0·2 μm pore size may pose a problem as the filters clog easily and less water can be processed (Turner *et al.* 2014a). This can be overcome by pre-filtering (Robson *et al.* 2016) and/or increasing the number of filter replicates. Future research is needed to identify optimal procedures for highly productive and/or turbid waters.

Conclusion

In conclusion, we recommend SX filters as an efficient capture method for aqueous eDNA sampling of macro-organisms. Preservation of SX in ethanol or Longmire's buffer immediately after filtration is recommended. Preserved SX capsules may be stored at RT for at least 2 weeks without significant degradation. Water samples can be quickly filtered and preserved on site requiring less equipment, easing transport. Therefore, SX capsules are logistically compatible with remote and harsh field conditions.

Authors' contributions

M.H. and J.S. conceived and designed initial experiment. All authors (except D.H.) contributed to final design and participated in 'sample collection/filtration day'. J.S. analysed data and drafted the manuscript. M.H. developed protocol for eDNA capture/extraction. J.S., M.H. and A.E. wrote the manuscript. A.E. and S.S.T.M. coordinated field experiment and contributed to extraction protocol. A.E., M.H., S.W.K., S.S.T.M., E.E.S. and M.S. extracted DNA. S.W.K. optimized qPCR protocol. S.W.K., M.H. and M.S. performed qPCR assays. All authors revised the manuscript. No conflict of interest exists.

Acknowledgements

We thank Philip Francis Thomsen (PFT), Ian Eirød and Peter Rask Møller for participating in 'experimental planning' and 'collection/filtration' day; PFT for providing qPCR primers/probes and collaborating on earlier SX extraction protocols; Eske Willerslev for laboratory facilities, funding and comments on earlier drafts.

Data accessibility

Data are deposited in the Dryad Data Repository http://dx.doi.org/10.5061/dryad.p2q4r (Spens *et al.* 2016).

References

Bergman, P.S., Schumer, G., Blankenship, S. & Campbell, E. (2016) Detection of adult green sturgeon using environmental DNA analysis. *PLoS One*, **11**, e0153500.

Bustin, S.A., Benes, V., Garson, J.A. *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, **55**, 611–622.

Chestnut, T., Anderson, C., Popa, R., Blaustein, A.R., Voytek, M., Olson, D.H. & Kirshtein, J. (2014) Heterogeneous occupancy and density estimates of the pathogenic fungus *Batrachochytrium dendrobatidis* in waters of North America. *PLoS One*, **9**, e106790.

Davy, C.M., Kidd, A.G. & Wilson, C.C. (2015) Development and validation of environmental DNA (eDNA) markers for detection of freshwater turtles. *PLoS One*, **10**, e0130965.

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.

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.

Eichmiller, J.J., Miller, L.M. & Sorensen, P.W. (2016) Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Molecular Ecology Resources*, **16**, 56–68.

Ellison, S.L., English, C.A., Burns, M.J. & Keer, J.T. (2006) Routes to improving the reliability of low level DNA analysis using real-time PCR. *BMC Biotechnology*, **6**, 33.

Ficetola, G.F., Miaud, C., Pompanon, F. & Taberlet, P. (2008) Species detection using environmental DNA from water samples. *Biology letters*, **4**, 423–425.

Gaillard, C. & Strauss, F. (1998) Avoiding adsorption of DNA to polypropylene tubes and denaturation of short DNA fragments. *Technical Tips Online*, **3**, 63–65.

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.

Höss, M., Kohn, M., Paabo, S., Knauer, F. & Schroder, W. (1992) Excrement analysis by PCR. *Nature*, **359**, 199.

IBM Corp. (2013) *IBM SPSS Statistics for Windows*, Version 22.0. IBM, Armonk, NY, USA.

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.

Keskin, E. (2014) Detection of invasive freshwater fish species using environmental DNA survey. *Biochemical Systematics and Ecology*, **56**, 68–74.

Liang, Z. & Keeley, A. (2013) Filtration recovery of extracellular DNA from environmental water samples. *Environmental Science & Technology*, **47**, 9324–9331.

Longmire, J.L., Maltbie, M. & Baker, R.J. (1997) Use of "lysis buffer" in DNA isolation and its implication for museum collections. *Occasional Papers the Museum Texas Tech University*, **163**, 1–3.

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 Research*, **39**, 541–548.

McKee, A.M., Calhoun, D.L., Barichivich, W.J., Spear, S.F., Goldberg, C.S. & Glenn, T.C. (2015) Assessment of environmental DNA for detecting presence of imperiled aquatic amphibian species in isolated wetlands. *Journal of Fish and Wildlife Management*, **6**, 498–510.

Minamoto, T., Naka, T., Moji, K. & Maruyama, A. (2016) Techniques for the practical collection of environmental DNA: filter selection, preservation, and extraction. *Limnology*, **17**, 23–32.

Piaggio, A.J., Engeman, R.M., Hopken, M.W., Humphrey, J.S., Keacher, K.L., Bruce, W.E. & Avery, M.L. (2014) Detecting an elusive invasive species: a diagnostic PCR to detect Burmese python in Florida waters and an assessment

of persistence of environmental DNA. *Molecular Ecology Resources*, **14**, 374–380.

Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R.M. & Gough, K.C. (2014) 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.

Renshaw, M.A., Olds, B.P., Jerde, C.L., McVeigh, M.M. & Lodge, D.M. (2015) The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol–chloroform–isoamyl alcohol DNA extraction. *Molecular Ecology Resources*, **15**, 168–176.

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.

Ross, K.S., Haites, N.E. & Kelly, K.F. (1990) Repeated freezing and thawing of peripheral blood and DNA in suspension: effects on DNA yield and integrity. *Journal of Medical Genetics*, **27**, 569–570.

Simmons, M., Tucker, A., Chadderton, W.L., Jerde, C.L. & Mahon, A.R. (2016) Active and passive environmental DNA surveillance of aquatic invasive species. *Canadian Journal of Fisheries and Aquatic Sciences*, **73**, 76–83.

Smart, A.S., Tingley, R., Weeks, A.R., van Rooyen, A.R. & McCarthy, M.A. (2015) Environmental DNA sampling is more sensitive than a traditional survey technique for detecting an aquatic invader. *Ecological applications*, **25**, 1944–1952.

Spens, J., Evans, A.R., Halfmaerten, D., Knudsen, S.W., Sengupta, M.E., Mak, S.S.T., Sigsgaard, E.E. & Hellström, M. (2016) Data from: Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Dryad Digital Repository*, http://dx.doi.org/10.5061/dryad.p2q4r

Sterivex™ (2013) *User Guide Sterivex™-GP Sterile Vented Filter Unit, 0.22 μm Single Use Only*. EMD Millipore Corporation, Billerica, MA, USA.

Takahara, T., Minamoto, T. & Doi, H. (2015) Effects of sample processing on the detection rate of environmental DNA from the Common Carp (*Cyprinus carpio*). *Biological Conservation*, **183**, 64–69.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., Orlando, L. & Willerslev, E. (2012) Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, **21**, 2565–2573.

Turner, C.R., Barnes, M.A., Xu, C.C.Y., Jones, S.E., Jerde, C.L. & Lodge, D.M. (2014a) Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution*, **5**, 676–684.

Turner, C.R., Miller, D.J., Coyne, K.J. & Corush, J. (2014b) Improved methods for capture, extraction, and quantitative assay of environmental DNA from Asian bigheaded carp (*Hypophthalmichthys* spp.). *PLoS One*, **9**, e114329.

Valentini, A., Taberlet, P., Miaud, C. *et al*. (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, **25**, 929–942.

Wegleitner, B., Jerde, C., Tucker, A., Chadderton, W.L. & Mahon, A. (2015) Long duration, room temperature preservation of filtered eDNA samples. *Conservation Genetics Resources*, **7**, 789–791.

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.

Willerslev, E., Hansen, A.J., Binladen, J. *et al*. (2003) Diverse plant and animal genetic records from holocene and pleistocene sediments. *Science*, **300**, 791–795.

Wilson, C.C., Wozney, K.M. & Smith, C.M. (2016) Recognizing false positives: synthetic oligonucleotide controls for environmental DNA surveillance. *Methods in Ecology and Evolution*, **7**, 23–29.

Zhan, A.B., Hulak, M., Sylvester, F. *et al*. (2013) High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution*, **4**, 558–565.

Received 29 May 2016; accepted 21 September 2016
Handling Editor: Douglas Yu

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Fig. S1. Flow chart illustrating the different capture and storage treatments.

Appendix S1. eDNA extraction protocol.

Appendix S2. Water quality in Gentofte lake.

Table S1. Empirical field-studies targeting macrobial eDNA in aquatic ecosystems with water sampling, January 2005 to March 2015.

Table S2. Empirical field-studies targeting macrobial eDNA in aquatic ecosystems with water sampling, published after the current study was initiated in March 2015.

Table S3. Primers and probes used in this study.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Fig. S1. Flow chart illustrating the different capture and storage treatments.

Appendix S1. eDNA extraction protocol.

Appendix S2. Water quality in Gentofte lake.

Table S1. Empirical field-studies targeting macrobial eDNA in aquatic ecosystems with water sampling, January 2005 to March 2015.

Table S2. Empirical field-studies targeting macrobial eDNA in aquatic ecosystems with water sampling, published after the current study was initiated in March 2015.

Table S3. Primers and probes used in this study.

Supplement

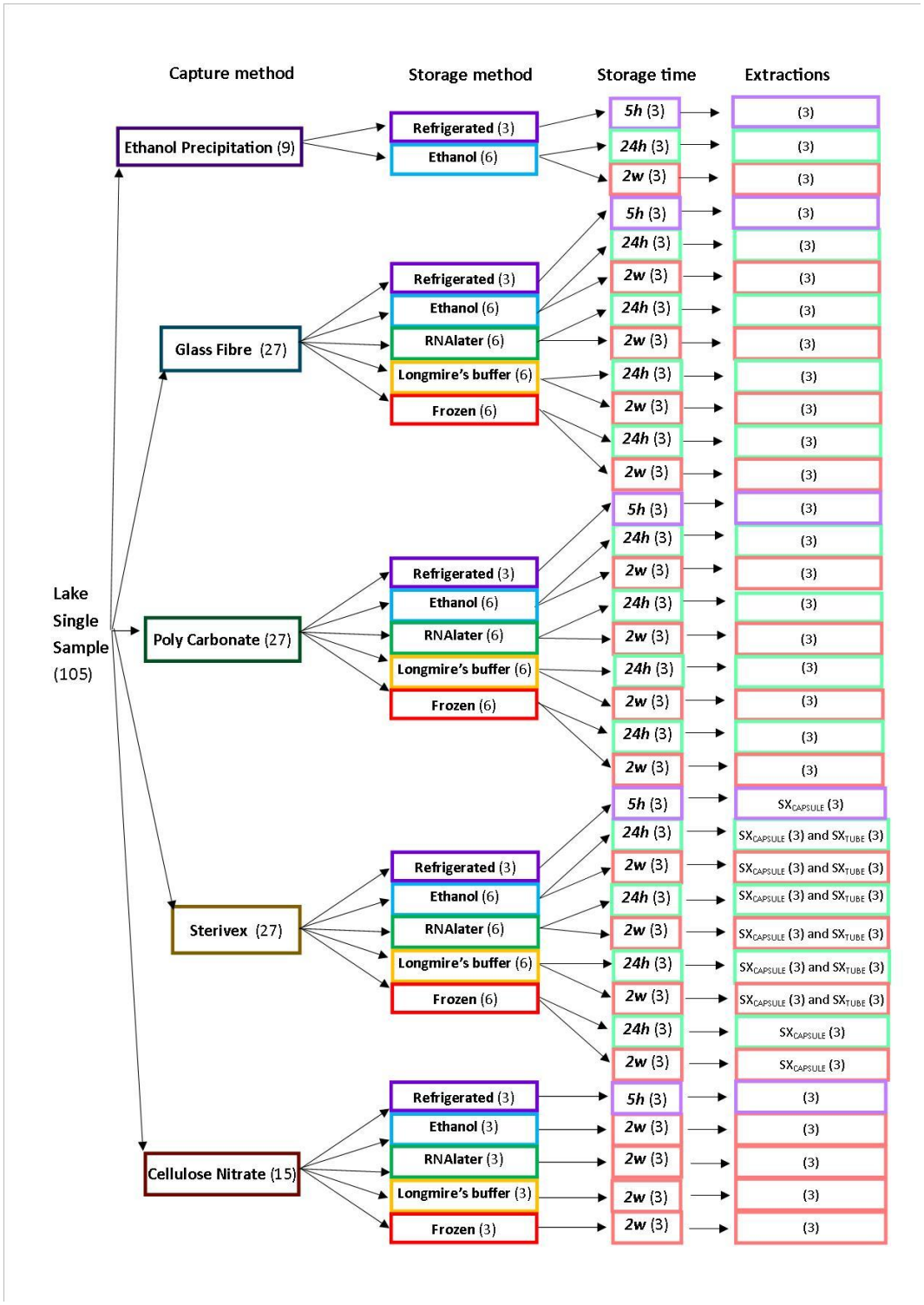


Figure S1. Flow chart illustrating the different capture and storage treatments Number of replicates in brackets. The abbreviations are outlined in the main text.

Appendix S1. eDNA extraction protocol

-modified from the DNeasy® blood & tissue kit (QIAGEN, Stockach, Germany) handbook pp.28-30, ver. 07/2006 (Qiagen®).

1. Before extraction:

Carefully wipe the outer surfaces of all the collection tubes and filter capsules with 5% bleach using clean tissue paper. Dry and wipe with 70% Ethanol using tissue paper.

2. Sample preparation until lysis buffer addition:

2 a *SX filters without preservation buffer (Frozen or Refrigerated).*

Proceed to **4b**.

2 b *SX filters with preservation buffer (ethanol, Longmire's or RNA later)*

2 b.1 SX_{TUBE}

Transfer the buffer from the filter capsule and into a 2mL sterile LoBind tube, with a 5 mL Luer-Lock syringe. Be careful not to apply too much pressure. Go to step **3a**.The extractions from the buffers are hereafter referred to as SX_{TUBE}.

2 b.2 $SX_{CAPSULE}$

After removal of buffers (step 2b.1) consider the SX capsules as test tubes. The filter will remain intact in the capsules to avoid loss of DNA and contamination risk by unnecessary handling. Remove the inlet and the outlet caps. In a tube rack placed inside a fume hood, dry the filters by placing them vertically with the 'inlet end' facing down. Let them blot on clean laboratory tissue paper placed underneath the rack. After drying follow the exact procedure of step **4b**. The extractions from the filters are hereafter referred to as $SX_{CAPSULE}$.

2 c *GF-, CN- and PC- Filter Samples without preservation buffers*

To each sample add 800 µL working solution as outlined in *step 4*. Vortex for 15s. Proceed to step 3.

2 d *GF-, CN- and PC- Filter Samples with preservation buffers (Ethanol, Longmire's or RNA later)*

Remove filter from the 5 mL LoBind tube (Eppendorf AG, Hamburg, Germany) with sterile forceps. Squeeze the liquid into the tube the filters were stored in. Dry filters by on the edge of a clean tube in the fume hood. Meanwhile for the buffers go to step 3b.:

2 e *Ethanol Precipitation (EP) Samples:*

If samples have been at RT before extraction, store in -20°C for 24 hours to enable efficient precipitation. Centrifuge the 50 mL Falcon tubes (Thermo Fisher Scientific, Walthma, MA, USA) for 30-45 minutes at 5,000 ∗ g (7,400 rpm using a 50 mL rotor). Discard supernatant and let pellet dry. Immediately before the next step; Prepare a lysis working-solution (reagents provided with the extraction kit) containing 720 µL ATL and 80 µL proteinase K/sample. To each sample add 800 µL working solution. Vortex for 15 s. Transfer to 2 mL LoBind tube (Eppendorf AG, Hamburg, Germany). Proceed to step 3

3. Centrifugation step:

3 a SX_{TUBE}

Spin at 6,000 ∗ g (8,000 rpm) for 30-45 min in a micro-centrifuge 24 ∗ 2mL. Discard liquid. Let pellet dry. Go to step 4a.

3 b *GF-, CN- and PC- Filter Samples with preservation buffers (Ethanol, Longmire's or RNA later)*

Spin down the buffers (preferably at 4°C) at 6,000 ∗ g (8,000 rpm) using a 5 mL rotor for 30-45 minutes. Discard supernatant and let pellet dry. Place dried filter from 2d in the corresponding 'pellet-tube' Proceed to step 4c

3 c *Ethanol Precipitation (EP) Samples*

Spin at 5,000 ∗ g 30-45 min using a 50 mL rotor. Discard liquid, let pellet dry, Proceed to step 4d.

4. Addition of lysis buffer:

Immediately before the lysis step; Make a premix of Lysis working solution by adding 720 µL ATL buffer and 80 µL proteinase K per sample provided by the kit. For SX_{TUBE} mix 180 µL ATL buffer and 20 µL proteinase K per sample.

4 a *SX_{TUBE}*

Dissolve the dried pellet by using 200 µL working solution (step 4)/ sample. Close tube and seal with parafilm. Vortex for 15 s and proceed to step 5.

4 b *$SX_{CAPSULE}$*

Keep the outlet end closed with the outlet cap (MOBIO 14600-50-NF-OC, QIAGEN, Stockach, Germany). Carefully add 800 µL Lysis working solution (step 4) to the filter by using a 1,000 µL pipet and sterile filter tips. Pipet the solution between the outside of the filter and the capsule walls Close with an inlet cap (MOBIO 14600-50-NF-IC, QIAGEN, Stockach, Germany), seal with parafilm. Handshake vigorously for a few seconds. Proceed to step 5.

4 c *GF-, CN- and PC- Filter Samples without and with preservation buffers (Ethanol, Longmire's or RNA later)*

For samples from step 3d dissolve pellet in an aliquot of working solution (step 4). For all samples in this step: Add 800 µL Lysis working solution/ sample. Close tube and seal with parafilm Vortex for 15 s and go to step **5**.

4 d *Ethanol Precipitation (EP) Samples*

Add 800 µL Lysis working solution (step4)/sample. Dissolve pellet and transfer to a 2 mL LoBind Eppendorf tube.

5. Incubate, while rotating, at 56°C for 24 hours.

6. Handshake SX filter capsules vigorously 5 times. **Vortex** the other samples for 15 s**.**

SX_{TUBE} (4a) samples and EP (*4d*) samples - proceed to step **8**.

GF, CN, PC (*4c*) proceed to step **7a**.

$SX_{CAPSULE}$ samples (4b) proceed to **7b**

7. Transfer:

7 a Measure the volume. Vortex for a few seconds. Spin down for 2 seconds to seed out excess debris. Transfer ALL liquid to 5 mL LoBind tube. Go to step **8**.

7 b Remove ALL the liquid from inlet end of capsule by using a Luer Lock syringe. Measure the volume, transfer to 5 mL LoBind tube. Vortex for a few seconds. Spin down for 2 seconds to seed out excess debris. Go to step **8**.

8. **Add Buffer** AL and ice cold molecular grade 99% ethanol (Thermo Fisher Scientific, Waltham, MA, USA) to the sample in equal volumes. Sample:Buffer:Ethanol = 1:1:1. Note: AL and ethanol can be premixed.
9. **Vortex** vigorously.
10. **Pipet** the mixture (max 650 µL at a time) into a DNeasy Mini Spin column in a 2 mL collection tube provided in the kit.
11. **Spin** in micro-centrifuge preferably at 4°C at 6000 ∗ g (8000 rpm for rotor max capacity 24 ∗ 1.5-2 mL tubes) 1 min.
12. **Discard** flow through.
13. **Repeat** steps 10-12 until all sample is filtered through DNeasy Mini spin column
14. Place the **DNeasy Mini spin column** in a new 2 ml collection tube (provided), add 500 µl Buffer AW1, and centrifuge for 1 min at 6000 ∗ g (8,000 rpm). Discard flow-through and collection tube. (QiaGen protocol)
15. Place the **DNeasy Mini spin column** in a new 2 ml collection tube (provided), add 500 µl Buffer AW2, and centrifuge for 3 min at 20,000 ∗ g (14,000 rpm) to dry the DNeasy membrane. Discard flow-through and collection tube. Place spin column in a new collection tube, centrifuge 1 min at 17,000 ∗ g (13,000 rpm).
16. **Transfer** spin column to a new 1.5 or 2 mL DNA LoBind tube with caps removed.

17. Place tubes with spin columns, four at a time, on a 70°C **heating plate**, add 100 µl 70°C Buffer TE (pH 8.0) to the membrane, immediately transfer spin column with filter to RT.
18. **Incubate** at RT for 10 min.
19. **Centrifuge** for 1 min at 6,000 ∗ g (8,000 rpm)
20. **Re-elute DNA** from DNA LoBind tube. (Apply eluate back on spin column on heating plate).
21. **Incubate** at RT for 10 min.
22. **Centrifuge** for 1 min at 6,000 ∗ g (8,000 rpm)
23. **Discard** the spin column.
24. **Transfer** DNA to pre-marked DNA LoBind tube with lid intact.
25. **Aliquot** 2 µL in a separate tube for DNA measurement.
26. **Store** at -20°C or at -80°C.

Appendix S2. Water quality in Gentofte lake

Water sample March 2015, Water Colour = 20 mg L^{-1} (i.e. clear water) was measured with a spectrometer in a 5 cm cuvette at a wavelength of 420 nm according to SS EN ISO 7887. Swedish standard methods for water quality are available from the Swedish Standards Institute, 118 80 Stockholm, Sweden (e-mail: info@sis.se).

Gentofte Lake (26 hectares) is designated as an EU Natura 2000 protected area representative of the habitat-type H3140 'Hard oligo-mesotrophic waters' in the Habitats Directive - Annex 1. European Union Council Directive 92/43/EEC on the Conservation of natural habitats and of wild fauna and flora.

Table S1. Empirical field-studies targeting macrobial eDNA in aquatic ecosystems with water sampling, January 2005 to March 2015

AUTHORS	CAPTURE	FILTER Poresize(µm)	FILTER Storage temperature (°C)	FILTER Storage medium	PCR TYPE	cycles	Total eDNA measured by	Total eDNA (ng uL^{-1})	SPECIES
Biggs et al. 2015	EP	---	-20		qPCR	55			Amphibian
Deiner & Altermatt 2014	GF	0.22	---	---	PCR	50			Arthropoda, Molluscs
Deiner et al. 2015	GF & EP	0.7	---	---	PCR, NGS	35	Qubit	0.24 & 0.43	Arthropoda, Molluscs
Dejean et al. 2011	EP	---	-20		PCR	55			Fish & Amphibian
Dejean et al. 2012	EP	---	-20		PCR	55			Amphibian
Díaz-Ferguson et al. 2014	CN	0.45	---	---	qPCR	40	NanoDrop	14.5 to 141	Fish
Egan et al. 2013	PC	20	-20	Dry	PCR & LTS	30	Nanodrop	116	Molluscs
Egan et al. 2015	PC	1.2	---	---	PCR & LTS	30	Qubit	5.54	Molluscs
Eichmiller et al. 2014	GF	1.5	-80	Dry	qPCR	40			Fish
Farrington et al. 2015	Centrifuge	---	---	---	PCR or qPCR	45 or 40			Fish
Ficetola et al. 2008	EP	---	-20		PCR	55			Amphibian
Foote et al. 2012	EP	---	-20		qPCR	55			Mammal
Fukumoto et al. 2015	GF	0.7	-25	Dry	qPCR	55			Amphibian
Goldberg et al. 2011	CN	0.45	RT	Ethanol & Dry	PCR	50, 55			Amphibians
Goldberg et al. 2013	CN or CEM	0.45	RT	Ethanol	qPCR	50			Molluscs
Huver et al. 2015	CN	3.0	-20	water	qPCR & PCR	40			Trematode
Jane et al. 2015	GF	1.5	-20 (-70)	Dry	qPCR	45			Fish
Janosik & Johnston 2015	GF	1.5	-20	Dry	PCR	35			Fish
Jerde et al. 2011	GF	1.5	-20	Dry	PCR	45			Fish
Jerde et al. 2013	GF	1.5	-20	Dry	PCR	45			Fish
Kortbaoui et al. 2009	CN	0.22 or 0.45	---	---	PCR (nested)	35	NanoDrop	max 0.1–0.2	Mammals, Bird
Keskin 2014	SX	0.22	-20	Dry	PCR	55			Fish
Laramie et al. 2015	CN	0.45	4	Ethanol	qPCR	50			Fish
Mahon et al. 2013	GF	1.5	-20	Dry	PCR	30			Fish
McKee et al. 2015	CN	0.45	RT	Ethanol 95%	qPCR	50	NanoDrop	0.015	Amphibian
Martellini et al. 2005	Centrifuge	---	RT	---	PCR (nested)	35			Mammals
Minamoto et al. 2012	ULTRAFILTER etc.	---	---	---	PCR	35			Fish
Moyer et al. 2014	CN	0.45	Frozen	Dry	qPCR	35			Fish
Mächler et al. 2014	GF	0.7	-20	Dry	PCR	50	Qubit	1 to 108	Arthropoda, Molluscs
Nathan et al. 2014a	GF	1.5	-20	Dry	PCR	30, 35			Fish
Olson et al. 2012	GF	1.5	-20	Dry	PCR	35			Amphibian
Piaggio et al. 2014	EP or GF	0.75	---	---	PCR	55			Reptile
Pilliod et al. 2013	CN	0.45	RT	Ethanol	qPCR	50			Amphibian
Pilliod et al. 2014	CN	0.45	RT	Ethanol	qPCR	50			Amphibian
Rees et al. 2014	EP	---	-20		qPCR	55			Amphibian
Santas et al. 2013	CELLULOSE	0.45	-20	PowerWater bead	PCR	55			Amphibian
Sigsgaard et al. 2015	EP	---	-20		qPCR	50			Fish
Spear et al. 2015	CN	0.45	RT	Ethanol	qPCR	50			Amphibian
Takahara et al. 2012	CA & ULTRAFILTER	3.0	-18 and -25	Dry	qPCR	40			Fish
Takahara et al. 2013	CA & ULTRAFILTER	3.0	-25	Dry	qPCR	55			Fish
Takahara et al. 2015	CA & ULTRAFILTER	3.0	-30	Ethanol	qPCR	40, 55			Fish
Thomsen et al. 2012a	EP	---	-20		454 & qPCR	45 & 55			Fish, Crustacean, Insect, Amphibians, Mammal
Thomsen et al. 2012b	NYLON	0.45	---	---	454 & qPCR	50 & 55			Fish
Tre´guier et al. 2014	EP	---	-20		qPCR	45			Crustacea
Turner et al. 2014a	EMD Nylon & PC	MULTI	-20	CTAB	qPCR	55	Qubit	0.004 to 0.008	Fish
Turner et al. 2014b	GF or PC	1.5 or 10	-20	Dry	PCR or qPCR	45 or 55			Fish
Turner et al. 2015	EP	---	-20 and -80		qPCR	55			Fish
Wilcox et al. 2013	GF	1.5	on ice	Dry	qPCR	45			Fish
Vuong et al. 2013	Sterile Millipore	0.45	-80	Dry	qPCR	40			Birds, Mammals
Coauthors in Spens et al. M.H., D.H., M.E.S. & S.W.K. Unpublished data	SX	0.22	---	---	qPCR, NGS	50	Qubit	7 to 269	Fish, Amphibians, Birds, Molluscs, Mammals, Arthropods, Trematodes

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.

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.

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.

Díaz-Ferguson, E., Herod, J., Galvez, J. & Moyer, G. (2014) Development of molecular markers for eDNA detection of the invasive African jewelfish (*Hemichromis letourneuxi*): a new tool for monitoring aquatic invasive species in National Wildlife Refuges. *Management of Biological Invasions,* **5,** 121-131.

Egan, S.P., Barnes, M.A., Hwang, C.-T., Mahon, A.R., Feder, J.L., Ruggiero, S.T., Tanner, C.E. & Lodge, D.M. (2013) Rapid invasive species detection by combining environmental DNA with light transmission spectroscopy. *Conservation Letters,* **6,** 402-409.

Egan, S.P., Grey, E., Olds, B., Feder, J.L., Ruggiero, S.T., Tanner, C.E. & Lodge, D.M. (2015) Rapid molecular detection of invasive species in ballast and harbor water by integrating environmental DNA and light transmission spectroscopy. *Environmental Science & Technology,* **49,** 4113-4121.

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.

Farrington, H.L., Edwards, C.E., Guan, X., Carr, M.R., Baerwaldt, K. & Lance, R.F. (2015) Mitochondrial genome sequencing and development of genetic markers for the detection of DNA of invasive bighead and silver carp (*Hypophthalmichthys nobilis* and *H. molitrix*) in environmental water samples from the United States. *PLoS One,* **10,** e0117803.

Ficetola, G.F., Miaud, C., Pompanon, F. & Taberlet, P. (2008) Species detection using environmental DNA from water samples. *Biology letters,* **4,** 423-425.

Foote, A.D., Thomsen, P.F., Sveegaard, S., Wahlberg, M., Kielgast, J., Kyhn, L.A., Salling, A.B., Galatius, A., Orlando, L. & Gilbert, M.T.P. (2012) Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. *PLoS One,* **7,** e41781.

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.

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., Sepulveda, A., Ray, A., Baumgardt, J. & Waits, L.P. (2013) Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshwater Science,* **32,** 792-800.

Huver, J.R., Koprivnikar, J., Johnson, P.T.J. & Whyard, S. (2015) Development and application of an eDNA method to detect and quantify a pathogenic parasite in aquatic ecosystems. *Ecological applications,* **25,** 991-1002.

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.

Janosik, A.M. & Johnston, C.E. (2015) Environmental DNA as an effective tool for detection of imperiled fishes. *Environmental Biology of Fishes,* **98,** 1889-1893.

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., 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.

Keskin, E. (2014) Detection of invasive freshwater fish species using environmental DNA survey. *Biochemical Systematics and Ecology,* **56,** 68-74.

Kortbaoui, R., Locas, A., Imbeau, M., Payment, P. & Villemur, R. (2009) Universal mitochondrial PCR combined with species-specific dot-blot assay as a source-tracking method of human, bovine, chicken, ovine, and porcine in fecal-contaminated surface water. *Water Research,* **43,** 2002-2010.

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.

Mahon, A.R., Jerde, C.L., Galaska, M., Bergner, J.L., Chadderton, W.L., Lodge, D.M., Hunter, M.E. & Nico, L.G. (2013) Validation of eDNA surveillance sensitivity for detection of Asian carps in controlled and field experiments. *PLoS One,* **8,** e58316.

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 Research,* **39,** 541-548.

McKee, A.M., Spear, S.F. & Pierson, T.W. (2015) The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biological Conservation,* **183,** 70-76.

Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M. & Kawabata, Z.i. (2012) Surveillance of fish species composition using environmental DNA. *Limnology*, **13,** 193–197.

Moyer, G.R., Díaz-Ferguson, E., Hill, J.E. & Shea, C. (2014) Assessing environmental DNA detection in controlled lentic systems. *PLoS One,* **9,** e103767.

Mächler, E., Deiner, K., Steinmann, P. & Altermatt, F. (2014) Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. *Freshwater Science,* **33,** 1174-1183.

Nathan, L.R., Jerde, C.L., Budny, M.L. & Mahon, A.R. (2014) The use of environmental DNA in invasive species surveillance of the Great Lakes commercial bait trade. *Conservation Biology,* **29,** 430-439.

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.

Piaggio, A.J., Engeman, R.M., Hopken, M.W., Humphrey, J.S., Keacher, K.L., Bruce, W.E. & Avery, M.L. (2014) Detecting an elusive invasive species: a diagnostic PCR to detect Burmese python in Florida waters and an assessment of persistence of environmental DNA. *Molecular Ecology Resources,* **14,** 374-380.

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.

Rees, H.C., Bishop, K., Middleditch, D.J., Patmore, J.R.M., Maddison, B.C. & Gough, K.C. (2014) The application of eDNA for monitoring of the Great Crested Newt in the UK. *Ecology and Evolution,* **4,** 4023-4032.

Santas, A.J., Persaud, T., Wolfe, B.A. & Bauman, J.M. (2013) Noninvasive Method for a Statewide Survey of Eastern Hellbenders *Cryptobranchus alleganiensis* Using Environmental DNA. *International Journal of Zoology,* **2013,** 174056.

Sigsgaard, E.E., Carl, H., Møller, 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.

Spear, S.F., Groves, J.D., Williams, L.A. & Waits, L.P. (2015) Using environmental DNA methods to improve detectability in a hellbender (*Cryptobranchus alleganiensis*) monitoring program. *Biological Conservation,* **183,** 38-45.

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.

Takahara, T., Minamoto, T. & Doi, H. (2015) Effects of sample processing on the detection rate of environmental DNA from the Common Carp (*Cyprinus carpio*). *Biological Conservation,* **183,** 64-69.

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H. & Kawabata, Z.i. (2012) Estimation of fish biomass using environmental DNA. *PLoS One,* **7,** e35868.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Møller, P.R., Rasmussen, M. & Willerslev, E. (2012) Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS One,* **7,** e41732.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., Orlando, L. & Willerslev, E. (2012) Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology,* **21,** 2565-2573.

Tréguier, A., Paillisson, J.-M., Dejean, T., Valentini, A., Schlaepfer, M.A. & Roussel, J.-M. (2014) Environmental DNA surveillance for invertebrate species: advantages and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. *Journal of Applied Ecology,* **51,** 871-879.

Turner, C.R., Barnes, M.A., Xu, C.C.Y., Jones, S.E., Jerde, C.L. & Lodge, D.M. (2014) Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution,* **5,** 676-684.

Turner, C.R., Miller, D.J., Coyne, K.J. & Corush, J. (2014) Improved methods for capture, extraction, and quantitative assay of environmental DNA from Asian bigheaded carp (*Hypophthalmichthys* spp.). *PLoS One,* **9,** e114329.

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.

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.

Vuong, N.-M., Villemur, R., Payment, P., Brousseau, R., Topp, E. & Masson, L. (2013) Fecal source tracking in water using a mitochondrial DNA microarray. *Water Research,* **47,** 16-30.

Table S2. Empirical field-studies targeting macrobial eDNA in aquatic ecosystems with water sampling, published after the current study was initiated in March 2015

AUTHORS	CAPTURE	FILTER Poresize(µm)	FILTER Storage temperature (°C)	FILTER Storage medium	PCR TYPE	cycles	Total eDNA measured by	Total eDNA (ng uL^{-1})	SPECIES
Adrian-Kalchhauser & Burkhardt-Holm 2016	GF & EP	0.7	on ice	dry	PCR & tdPCR	15 + 35	---	---	Fish
Amberg et al. et al. 2015	GF	1.5	-80 then dry ice	dry	cPCR & qPCR	45	Nanodrop	60	Fish
Ardura et al. 2015	Nucleopore	0.12	---	96% ethanol	PCR	35	BioPhotometer	11 to 677	Molluscs
Bergman et al. 2016	SX	0.22	on ice then -20	---	qPCR	40	---	---	Fish
Boothroyd et al. 2016	GF	1 to 1.2	-80	dry	qPCR	40	---	---	Fish
Cannon et al. 2016	Centrifuge	---	---	---	PCR, NGS	50	---	---	Mammals, Amphibian, Birds, Arthropods, Fish, Plants, Bryophyta
Davy et al. 2015	GF	1.2	-80	dry	PCR & qPCR	35 & 40	---	---	Turtles
Doi et al. 2015	Cellulose Acetat	3	---	---	ddPCR & qPCR	45 & 55	---	---	Fish
Dougherty et al. 2016	CN, PCTE	1.2	refrigerator	Longmire's	qPCR	45	---	---	Crustacea
Furlan et al. 2016	GF	---	---	---	qPCR	55	---	---	Fish
Fujiwara et al. 2016	GF	---	---	---	qPCR	55	---	---	Plant
Furlan & Gleeson 2016	GF	1.2	-20	dry	qPCR	55	---	---	Fish
Gingera et al. 2016	GF	1.5	-30	ethanol	PCR	35	---	---	Fish
Gustavson et al. 2015	CN	0.45	-20	100% ethanol	qPCR	40	---	---	Fish
Hunter et al. 2015	CN	0.45	frozen	dry	qPCR	40	BIOTEK & Qubit	up to 309.5	Reptile
Hänfling et al. 2016	CN	0.45	-20	---	NGS	40	---	---	Fish
Koizumi et al. 2015	GF	0.7	-30	aluminium	PCR	40	---	---	Fish
Lacoursière-Roussel et al. 2016	GF	1.2	-20	dry	qPCR	---	---	---	Fish
McKee et al. 2015	CN	0.45	-20	95% ethanol	qPCR	50	---	---	Amphibian
McKelvey et al. 2016	GF	1.5	-20	---	qPCR	45	---	---	Fish
Minamoto et al. 2016	GF, PCTE, EP	0.2 to 3.0	---	---	qPCR	40	---	---	Fish
Miya et al. 2015	GF	0.7	-20	aluminium	NGS	35 + 12	---	---	Fish
Mächler et al. 2016	GF	0.7	on ice	tissue lysis buffer	PCR	50	Qubit	0.15 to 3.7	Mollusca, Insecta, Crustacea
Newton et al. 2016	PVDF, EP	0.45	-20	ethanol	PCR	55	---	---	Plant
Pierson et al. 2016	CN	0.45	-20	95% ethanol	qPCR	50	---	---	Amphibian
Piggott 2016	CN, EP	0.45	-20	ethanol	PCR & qPCR	45	Qubit	<20	Fish
Port et al. 2016	PVDF	0.22	-80	---	PCR, NGS	40	---	---	Fish, Birds, Mammals
Robson et al. 2016	PC, Nylon net	3, 10, 20	-20	dry	qPCR	40	---	---	Fish
Schmelzle & Kinziger 2016	PC	3	-20	dry	qPCR	55	---	---	Fish
Secondi et al. 2016	EP	---	-20	ethanol	qPCR	55	---	---	Amphibian
Shaw et al. 2016	CN	0.45	immediate extraction	no storage	NGS	35	---	---	Fish
Simmons et al. 2016	GF	1.5	-20	---	PCR, ddPCR,NGS	35, 40, 45	---	---	Fish
Smart et al. 2015	CN	0.45	4C	dry sterile	qPCR	50	---	---	Amphibian
Stoeckle et al. 2016	GF	0.22	-80	dry	qPCR nested	40	---	---	Mollusca
Uchii et al. 2016	GF	0.7	-20	---	qPCR	40 &50	---	---	Fish
Valentini et al. 2016	Envirochek HV	1	4 then -20	---	NGS	50	---	---	Fish, Amphibians
Wilcox et al. 2016	GF	1.5	ambient then -20	silica desiccant	qPCR	45	---	---	Fish
Yamamoto et al. 2016	GF	0.7	frozen	---	qPCR	55	---	---	Fish
Yamanaka & Minamoto 2016	GF	0.7	-20	dry	qPCR	55	---	---	Fish

Adrian-Kalchhauser, I. & Burkhardt-Holm, P. (2016) An eDNA assay to monitor a globally invasive fish species from flowing freshwater. *PLoS One,* **11,** e0147558.

Amberg, J.J., McCalla, S.G., Monroe, E., Lance, R., Baerwaldt, K. & Gaikowski, M.P. (2015) Improving efficiency and reliability of environmental DNA analysis for silver carp. *Journal of Great Lakes Research,* **41,** 367-373.

Ardura, A., Zaiko, A., Martinez, J.L., Samulioviene, A., Semenova, A. & Garcia-Vazquez, E. (2015) eDNA and specific primers for early detection of invasive species – A case study on the bivalve *Rangia cuneata*, currently spreading in Europe. *Marine Environmental Research,* **112B,** 48–55.

Bergman, P.S., Schumer, G., Blankenship, S. & Campbell, E. (2016) Detection of Adult Green Sturgeon Using Environmental DNA Analysis. *PLoS One,* **11,** e0153500.

Boothroyd, M., Mandrak, N.E., Fox, M. & Wilson, C.C. (2016) Environmental DNA (eDNA) detection and habitat occupancy of threatened spotted gar (*Lepisosteus oculatus*). *Aquatic Conservation: Marine and Freshwater Ecosystems,* DOI: 10.1002/aqc.2617.

Cannon, M.V., Hester, J., Shalkhauser, A., Chan, E.R., Logue, K., Small, S.T. & Serre, D. (2016) In silico assessment of primers for eDNA studies using PrimerTree and application to characterize the biodiversity surrounding the Cuyahoga River. *Scientific Reports,* **6,** 11.

Davy, C.M., Kidd, A.G. & Wilson, C.C. (2015) Development and validation of environmental DNA (eDNA) markers for detection of freshwater turtles. *PLoS One,* **10,** e0130965.

Doi, H., Takahara, T., Minamoto, T., Matsuhashi, S., Uchii, K. & Yamanaka, H. (2015) Droplet digital PCR outperforms real-time PCR in the detection of environmental DNA from an invasive fish species. *Environmental Science & Technology,* **49,** 5601–5608.

Dougherty, M.M., Larson, E.R., Renshaw, M.A., Gantz, C.A., Egan, S.P., Erickson, D.M. & Lodge, D.M. (2016) Environmental DNA (eDNA) detects the invasive rusty crayfish *Orconectes rusticus* at low abundances. *Journal of Applied Ecology,* **53,** 722–732.

Fujiwara, A., Matsuhashi, S., Doi, H., Yamamoto, S. & Minamoto, T. (2016) Use of environmental DNA to survey the distribution of an invasive submerged plant in ponds. *Freshwater Science,* **35,** 748-754.

Furlan, E.M. & Gleeson, D. (2016) Environmental DNA detection of redfin perch, *Perca fluviatilis*. *Conservation Genetics Resources*, **8,** 115–118.

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.

Gingera, T.D., Steeves, T.B., Boguski, D.A., Whyard, S., Li, W. & Docker, M.F. (2016) Detection and identification of lampreys in Great Lakes streams using environmental DNA. *Journal of Great Lakes Research*, **42,** 649–659.

Gustavson, M.S., Collins, P.C., Finarelli, J.A., Egan, D., Conchúir, R.Ó., Wightman, G.D., King, J.J., Gauthier, D.T., Whelan, K., Carlsson, J.E.L. & Carlsson, J. (2015) An eDNA assay for Irish *Petromyzon marinus* and *Salmo trutta* and field validation in running water. *Journal of Fish Biology,* **87,** 1254-1262.

Hunter, M.E., Oyler-McCance, S.J., Dorazio, R.M., Fike, J.A., Smith, B.J., Hunter, C.T., Reed, R.N. & Hart, K.M. (2015) Environmental DNA (eDNA) sampling improves occurrence and detection estimates of invasive Burmese pythons. *PLoS One,* **10,** e0121655.

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.

Koizumi, N., Takahara, T., Minamoto, T., Doi, H., Mori, A., Watabe, K. & Takemura, T. (2015) Preliminary experiment for detection method of fish inhabiting agricultural drainage canals using environmental DNA. *Transactions of the Japanese Society of Irrigation, Drainage and Rural Engineering,* **83,** 47-48.

Lacoursière-Roussel, A., Côté, G., Leclerc, V. & Bernatchez, L. (2016) Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. *Journal of Applied Ecology,* **53,** 1148–1157.

McKee, A.M., Calhoun, D.L., Barichivich, W.J., Spear, S.F., Goldberg, C.S. & Glenn, T.C. (2015) Assessment of environmental DNA for detecting presence of imperiled aquatic amphibian species in isolated wetlands. *Journal of Fish and Wildlife Management,* **6,** 498-510.

McKelvey, K.S., Young, M.K., Knotek, W.L., Carim, K.J., Wilcox, T.M., Padgett-Stewart, T.M. & Schwartz, M.K. (2016) Sampling large geographic areas for rare species using environmental DNA: a study of bull trout *Salvelinus confluentus* occupancy in western Montana. *Journal of Fish Biology,* **88,** 1215–1222.

Minamoto, T., Naka, T., Moji, K. & Maruyama, A. (2016) Techniques for the practical collection of environmental DNA: filter selection, preservation, and extraction. *Limnology,* **17,** 23-32.

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,** 150088.

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.

Newton, J., Sepulveda, A., Sylvester, K. & Thum, R.A. (2016) Potential utility of environmental DNA for early detection of Eurasian watermilfoil (*Myriophyllum spicatum*). *Journal of Aquatic Plant Management,* **54,** 46-49.

Pierson, T.W., McKee, A.M., Spear, S.F., Maerz, J.C., Camp, C.D. & Glenn, T.C. (2016) Detection of an enigmatic plethodontid salamander using environmental DNA. *Copeia,* **104,** 78-82.

Piggott, M.P. (2016) Evaluating the effects of laboratory protocols on eDNA detection probability for an endangered freshwater fish. *Ecology and Evolution,* **6,** 2739–2750.

Port, J.A., O'Donnell, J.L., Romero-Maraccini, O.C., Leary, P.R., Litvin, S.Y., Nickols, K.J., Yamahara, K.M. & Kelly, R.P. (2016) Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology,* **25,** 527–541.

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.

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.

Secondi, J., Dejean, T., Valentini, A., Audebaud, B. & Miaud, C. (2016) Detection of a global aquatic invasive amphibian, *Xenopus laevis*, using environmental DNA. *Amphibia-Reptilia,* **37,** 131-136.

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.

Simmons, M., Tucker, A., Chadderton, W.L., Jerde, C.L. & Mahon, A.R. (2016) Active and passive environmental DNA surveillance of aquatic invasive species. *Canadian Journal of Fisheries and Aquatic Sciences,* **73,** 76-83.

Smart, A.S., Tingley, R., Weeks, A.R., van Rooyen, A.R. & McCarthy, M.A. (2015) Environmental DNA sampling is more sensitive than a traditional survey technique for detecting an aquatic invader. *Ecological applications,* **25,** 1944-1952.

Stoeckle, B.c., Kuehn, R. & Geist, J. (2016) Environmental DNA as a monitoring tool for the endangered freshwater pearl mussel (*Margaritifera margaritifera L.*): a substitute for classical monitoring approaches? *Aquatic Conservation: Marine and Freshwater Ecosystems,* DOI: 10.1002/aqc.2611.

Uchii, K., Doi, H. & Minamoto, T. (2016) A novel environmental DNA approach to quantify the cryptic invasion of non-native genotypes. *Molecular Ecology Resources,* **16,** 415-422.

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., Møller, P.R., Willerslev, E. & Dejean, T. (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology,* **25,** 929–942.

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.

Yamamoto, S., Minami, K., Fukaya, K., Takahashi, K., Sawada, H., Murakami, H., Tsuji, S., Hashizume, H., Kubonaga, S., Horiuchi, T., Hongo, M., Nishida, J., Okugawa, Y., Fujiwara, A., Fukuda, M., Hidaka, S., Suzuki, K.W., Miya, M., Araki, H., Yamanaka, H., Maruyama, A., Miyashita, K., Masuda, R., Minamoto, T. & Kondoh, M. (2016) Environmental DNA as a 'snapshot' of fish distribution: A case study of Japanese jack mackerel in Maizuru Bay, Sea of Japan. *PLoS One,* **11,** e0149786.

Yamanaka, H. & Minamoto, T. (2016) The use of environmental DNA of fishes as an efficient method of determining habitat connectivity. *Ecological Indicators,* **62,** 147-153.

Table S3. Primers and probes used in this study (courtesy of PFT). Fragment lengths are given in base-pairs including primers. The amplified gene is Cytb: Cytochrome b, Probes are Black Hole Quencher-1 (BHQ1) and have the modifications; 5': 6-Fam (D-L-Probe), 3': BHQ-1

Taxon	**Primer/probe**	**Sequence 5'-3' with modifications**	**Length (bp)**	**Gene**	**Optimal Primer-/ probe conc.** (nM) in 25 µL rxn
Pike *Perca fluviatilis*	PerfluCBL	ACGCTCGATTCCAAACAAAC	89	cyt b	200
	PerfluCBR	GTGTGAAGGATGGGGACAAC			1000
	PerfluCB.probe	FAM-GCCTTACTTGCCTCCATCCTGGTTC-BHQ1			300
Perch *Esox lucius*	EsolucCBL	GGGACGTTAACTACGGCTGA	84	cyt b	1200
	EsolucCBR	CGGGCGATGTGTATGTAAA			800
	EsolucCB.probe	FAM-CCGAAATATTCACGCTAACGGTGCA-BHQ1			300

Testing included *Umbra pygmea*, *Sander lucioperca*, *Abramis brama*, *P. fluviatilis*, *E. lucius*, *P. flavescens*, *Carassius carassius*, *Gymnocephalus cernua*, *Rutilus rutilus*, *Scardinius erythropthalmus*, *Thymallus thymallus*, *Anguilla anguilla* and *Salmo trutta*.

Appendix 3

How can we conserve the imperilled freshwater ecosystems of Southeast Asia?

Evans, A., von Rintelen, T., Rüber, L., Woodruff, D., Voris, H., Dudgeon, D., von Rintelen, K., Carvalho, G., Mather, P. B., Nugroho, E., Balke, M., Tan, H. H., Wowor, D., Kottelat, M., de Bruyn, M. (*in prep*). How can we conserve the imperilled freshwater ecosystems of Southeast Asia?

OXFORD UNIVERSITY PRESS | BioScience

How can we conserve the imperilled freshwater ecosystems of Southeast Asia?

Journal:	*BioScience*
Manuscript ID	Draft
Manuscript Type:	Forum
Date Submitted by the Author:	n/a
Complete List of Authors:	Evans, Alice; Bangor University, Molecular Ecology and Fisheries Genetics Laboratory Von Rintelen, Thomas; Museum fur Naturkunde - Leibniz-Institut fur Evolutions- und Biodiversitatsforschung, Leibniz Institute for Research on Evolution and Biodiversity Rüber, Lukas; NMBE, Ichthyology & Herpetology Woodruff, David; University of California, San Diego, Section of Ecology, Behavior, and Evolution; Voris, Harold; Field Museum of Natural History, Science and Education Dudgeon, David; University of Hong Kong, School of Biological Sciences Von Rintelen, Kristina; Museum fur Naturkunde - Leibniz-Institut fur Evolutions- und Biodiversitatsforschung, Leibniz Institute for Research on Evolution and Biodiversity Carvalho, Gary; Bangor University, Molecular Ecology and Fisheries Genetics Laboratory Mather, Peter; Queensland University of Technology Faculty of Science and Engineering, Earth, Environmental and Biological Sciences Nugroho, Estu; Research Institute for Freshwater Aquaculture, Aquaculture Balke, Michael; Zoologische Staatssammlung Munchen, Systematics Tan, Heok Hui; National University of Singapore, Lee Kong Chian Natural History Museum Wowor, Daisy; Research Center for Biology, Indonesian Institute of Sciences (LIPI), Division of Zoology Kottelat, Maurice; National University of Singapore, Lee Kong Chian Natural History Museum de Bruyn, Mark; The University of Sydney, School of Life and Environmental Sciences
Key words:	Conservation solutions, ecosystem threats, biodiversity, aquatic, policy
Abstract:	Southeast Asian freshwater ecosystems and their rich biota are under extreme and sustained threat, requiring immediate coordinated action by scientists, environmental advocacy groups, national governments, policy makers and the international community in order to conserve this unique biodiversity, as well as the critically important goods and services that freshwater ecosystems provide. Moreover, freshwaters have experienced over twice the rate of biodiversity loss compared to marine and terrestrial systems regionally. Major threats include water pollution, flow modification,

habitat degradation, over-exploitation, species invasions and climatic factors. Here, we propose seven key actions to conserve high biodiversity and endemism, as well as associated important goods and services provided by freshwater ecosystems.

FORUM: **How can we conserve the imperilled freshwater ecosystems of Southeast Asia?**

Authors: Alice Evans[1,2*], Thomas von Rintelen[3], Lukas Rüber[4,5], David Woodruff[6], Harold Voris[7], David Dudgeon[8], Kristina von Rintelen[3], Gary Carvalho[1], Peter B. Mather[9], Estu Nugroho[10], Michael Balke[11], Heok Hui Tan[12], Daisy Wowor[13], Maurice Kottelat[12], Mark de Bruyn[1,14,*]

Affiliations:

1. Molecular Ecology and Fisheries Genetics Laboratory, Environment Centre Wales, Bangor University, Bangor, UK

2. Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5–7, 1350 Copenhagen K, Denmark

3. Museum für Naturkunde, Leibniz Institute for Research on Evolution and Biodiversity, Humboldt University, Berlin, Germany

4. Naturhistorisches Museum der Burgergemeinde Bern, Bern, Switzerland

5. Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

6. Division of Biological Sciences, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0116, USA

7. Science and Education, Field Museum of Natural History, Chicago, IL, 60605, USA

8. School of Biological Sciences, The University of Hong Kong, Hong Kong SAR, China

9. Earth, Environmental and Biological Sciences, Bioscience, Queensland University of Technology, Brisbane, Australia

10. Research Institute for Freshwater Aquaculture, Jl.Sempur 1, Bogor, West Java, Indonesia

11. Zoologische Staatssammlung, Münchhausenstraße 21 81247 Munich, Germany

12. Lee Kong Chian Natural History Museum, National University of Singapore, 2 Conservatory Drive, Singapore 117377, Republic of Singapore

13. Division of Zoology, Research Center for Biology, Indonesian Institute of Sciences (LIPI), Jalan Raya Jakarta Bogor Km 46, Cibinong 16911, Indonesia.

14. School of Life and Environmental Sciences, University of Sydney, Sydney 2006, NSW, Australia.

* Authors for correspondence: E-mail: markus.debruyn@gmail.com; alice.evans@bangor.ac.uk

Abstract: Southeast Asian freshwater ecosystems and their rich biota are under extreme and sustained threat, requiring immediate coordinated action by scientists, environmental advocacy groups, national governments, policy makers and the international community in order to conserve this unique biodiversity, as well as the critically important goods and services that freshwater ecosystems provide. Moreover, freshwaters have experienced over twice the rate of biodiversity loss compared to marine and terrestrial systems regionally. Major threats include water pollution, flow modification, habitat degradation, over-exploitation, species invasions and climatic factors. Here, we propose seven key actions to conserve high biodiversity and endemism, as well as associated important goods and services provided by freshwater ecosystems.

Key words: Conservation solutions, ecosystem threats, biodiversity, aquatic, policy

Word count: 4,991

1. Introduction

When Alfred Russel Wallace sailed between the volcanic shores and hiked into the humid forests of the Malay Archipelago in the 1850s, the influence of man had done little to erode the ancient and flourishing biodiversity of Southeast Asia (SEA). Wallace's seminal book, *The Malay Archipelago* (1869), revealed the exceptional endemism of this region, and the stark division of species between the Asian and Australian continents on either side of what became known appropriately as Wallace's Line. More than one hundred and fifty years later, the Malay Archipelago encompasses most of modern day SEA, hosting four of the Earth's terrestrial biodiversity hotspots: Indo-Burma (Cambodia, Laos, Thailand, Vietnam and Myanmar), Sundaland (Brunei, Indonesia, Malaysia, Singapore), Philippines and Wallacea (Indonesia) (Myers et al. 2000). Unfortunately, the threats to its flora and fauna are greater than ever.

Conservation of freshwater ecosystems (FEs) is often overlooked, despite freshwater biodiversity declining faster than either terrestrial or marine biodiversity since 1970 (Dudgeon et al. 2006, Collen et al. 2014). Freshwater conservation strategies are of immediate critical importance in densely-populated regions such as SEA, where high rates of habitat loss and species extinction (Myers et al. 2000; Collen et al. 2014) (Figure 1) coincide with manifest risks to human water security (Vörösmarty et al. 2010) (Figure 2). As the global human population, sea-level, and temperatures rise, it is inevitable that threats to freshwater ecosystems will intensify. An increase in frequency of extreme weather events, combined with economic expansion and a tendency towards further expansion of coastal cities will exacerbate the effects of anthropogenic change in SEA, as evinced by events reported in the media since 2015. These include the forest fires that ravaged Indonesia, and were intensified by the drainage of Bornean and Sumatran wetlands; one of the most severe El Niño weather events recorded in 50 years that caused widespread drought; saline intrusion

that crept up the Vietnamese Mekong for the first time; and work on the final stage of the US$3.5bn Xayaburi Dam in Laos — the first dam on the mainstream of the lower Mekong River that sustains the world's largest freshwater capture fishery. It is therefore timely to review ongoing threats to freshwater ecosystems in SEA, and propose novel, realistic and effective solutions to protect them.

2. The status of freshwater ecosystems in SEA

Freshwater ecosystems occupy 0.01% of the water, and 0.8% of the surface of Earth, but contain ~126,000 plant and animal species (Balian et al. 2008), equivalent to ~ 9% of all described species. Almost double the rate of loss of biodiversity is observed in FEs compared to terrestrial and marine environments recorded between 1973 - 2000 (Collen et al. 2014). Although inventories are incomplete, globally and regionally 30-50% of freshwater fishes and amphibians are extinct or endangered (Dudgeon et al. 2006; Hails et al. 2008; Rowley et al. 2010) with freshwater fishes being the most threatened group of vertebrates (Reid et al. 2013). SEA ranks second globally (after the Amazon) for freshwater species richness, with the Mekong Basin and large parts of Malaysia and Indonesia considered noteworthy (Collen et al. 2014). It is the richest region on the planet for freshwater turtles (Buhlmann et al. 2009), and fish, crustacean, insect and molluscan diversity is particularly high (Balian et al. 2008; Kottelat 2013; De Grave et al. 2015) (Figure 3). Iconic taxa include both the world's heaviest freshwater fish (Mekong giant catfish, *Pangasianodon gigas*) and one of the smallest known vertebrates (a peat swamp forest dwarf minnow, *Paedocypris progenetica*) (Lévêque et al. 2008). This region also, unfortunately, has the highest number of threatened freshwater species on Earth (Figure 1; Collen et al. 2014).

Global extent of wetlands decreased by ~50% during the 20th Century (Hails et al. 2008), but the losses are certainly higher in SEA than globally (Rowley et al. 2010), and here

most remaining wetlands have been converted to rice paddy fields, reservoirs, canals or storm drains. There is reason to anticipate also that threat intensities could increase in future. For instance, SEA ecosystems have experienced repeated and significant geographic reductions associated with the periodic submergence of the Sunda Shelf during Pleistocene interglacial periods (Woodruff 2010). This repeated range compaction and subsequent expansion may account for the hyperdiverse communities we observe today. As current climate and geography are typical of only ~3% of the last 2.7 million years, the biota of SEA is currently in a refugial state, in which they occupy only 50-75% of their maximal Pleistocene extent (Woodruff 2010). Sea-level rise will impose additional threats through further reductions in land area and an associated increase in the refugial state for SEA taxa.

3. Threats to Southeast Asian freshwater ecosystems and ecosystem services

Human society has depended upon FEs for thousands of years, with the birth of early empires occurring in river valleys such as at Angkor Wat, and along the Nile, Indus and Ganges Rivers, as these sites provided fertile soils, plentiful fishing, timber, wild game, drinking water, irrigation and transport (Scott 1989). Amenities and processes provided by freshwater ecosystems and biodiversity such as extreme weather 'insurance', a repository of genetic information, and creation of clean water are frequently termed 'ecosystem services', a controversial classification assigning economic value to products and processes performed by an ecosystem. Some ecologists argue the prioritization of ecosystem services may actually be detrimental for conservation, as it takes little account of the innate value of biodiversity (Dudgeon 2014), whilst others believe the use of economic incentives is a necessary tool (Kareiva and Marvier 2012)

Freshwater habitats are structurally complex, with a range of spatial and temporal flows, rivers, lakes, surface-groundwater systems, lateral and longitudinal connectivity, patch

disturbance, and channel form which results in varied levels of biodiversity threat (Dudgeon et al. 2006; Vörösmarty et al. 2010). Accelerating anthropogenic impacts including flow modification, habitat destruction and degradation, pollution, overharvesting, introduced species and climate change now impact this complex balance, exacerbating the natural vulnerability of SEA's freshwater biota and compromising habitats (Dudgeon et al. 2006; Peh 2010; Collen et al. 2014, Welcomme et al. 2016). Their interaction has, and is likely to continue to, cause declines in fishery yields and abundance of large species (Welcomme et al. 2016). Animals which are particularly vulnerable are those that have low fecundity, late maturation of large size, strict habitat specialisation or narrow geographic ranges, and a reliance on annual flood-pulse cycles that often involves a breeding migration (Dudgeon 2011; Allen et al. 2012; Welcomme et al. 2016). This vulnerability of the biota is exacerbated by the natural features of FEs, which are prone to fragmentation, pollution, and establishment of invasive species (Dudgeon et al. 2006; Darwall et al. 2009).

The ~646 million humans in SEA require food, water, energy, consumables and living space, which threaten FEs through a range of interrelated activities such as oil and gas extraction, hydropower creation, agricultural development and urban expansion. SEA populations are forecast to reach 792 million by 2050 (Worldometers 2017), subjecting these services to ever-greater demand, and increasing habitat loss through river impoundment, urbanization, deforestation and land-use change.

3.1 Water pollution

Freshwater ecosystems are often 'receivers' of pollution as agricultural fertilizer, pesticides, industrial effluents, mining waste, domestic sewage, heavy metals and synthetic chemicals drain down the landscape into lakes, rivers and wetlands (Dudgeon et al. 2006; Cochard 2017). These pollutants create unsafe drinking water, hazards to aquatic biodiversity and

terrestrial wildlife, and oxygen depleted 'dead zones' as well as causing harm to humans. For example, Persistent Organic Pollutants (POPs) can cause damage to the nervous, immune, endocrine and reproductive systems, as well as birth defects and cancer in animals and humans (Triet et al. 2014). Many POPs banned under the Stockholm Convention are still used in countries of the Lower Mekong Basin; examples include Chlordane (agricultural insecticide), Endrin (agricultural pesticide) and Hexachlorobenzene (fungicide) (Triet et al. 2014). Bioaccumulation of such pollutants causes a direct threat to humans in SEA consuming certain animals such as freshwater fish and crustaceans (Allen et al. 2012, Greer et al. 2017). Excess nutrients from agricultural fertilizers can cause harmful algal blooms and a concomitant increase in cyanotoxins. A recent study showed two to fourteen times the tolerable daily intake value of cyanotoxins in tilapia fish from aquaculture farms in SEA, with potentially dangerous bioaccumulation effects for humans including hepatocellular damage, liver cancer, colorectal cancer and renal function (Greer et al. 2017). Healthy FEs containing key species act as natural pollutant filters (Chowdhury et al. 2016, Cochard 2017), which can be more effective economically than the construction and operation of water filtration plants (Collen et al. 2014). Intact palustrine wetlands can provide the ecosystems services of water pollution removal as well as swamp fisheries, biomass production, seasonal agriculture and wildlife conservation (Cochard 2017)

3.2 Flow modification

Flow modifications may occur through river impoundment (hydropower and reservoir dams), levees and channel modification, water diversions (for water extraction and agricultural irrigation), and surface and groundwater abstraction (Poff and Zimmerman 2009). These modifications are universal in FEs and their impacts on ecosystem services are most deleterious in locations with highly variable flow regimes, where humans have the greatest

need for flood protection or water storage (Vörösmarty et al. 2010; Dudgeon et al. 2006). For example, hydropower dams present a clean, renewable source of energy, but construction will alter natural flow regimes with consequences for water temperature, nutrient loads and sediment transport downstream, as well as contributing to terrestrial and aquatic species and habitat loss, reduction of fishery yield and deterring fish migration (Stone 2011; Winemiller et al. 2016, Welcomme et al. 2016). In SEA, 98 dams are planned for construction by 2030 in the Mekong basin alone, with an additional 371 dams already operational or under construction (see Figure 4). An increase of this magnitude would require a 19-63% expansion of agricultural land to preserve regional food security in the face of projected fishery loss (Winemiller et al. 2016). The high productivity of regional rivers is attributable to annual flood-pulse cycles driven by monsoonal rainfall, with many fish species in the middle Mekong migrating from the mainstream into tributaries and shallow flood plains to feed and breed (Dudgeon 2011; Kano et al. 2016). Prevention of fish migration is therefore one of the most destructive impacts of dams, and the provision of mitigation technology in the form of fishways, locks and lifts are thought to be largely ineffective on rivers such as the Mekong which involve tens of millions of individuals and over 50 species important to regional food security (Dugan et al. 2010).

3.3 Habitat degradation

Maintaining or enhancing habitat heterogeneity that creates niche opportunities is beneficial to both biodiversity and ecosystem functioning (Allen et al. 2012). Habitat degradation, and subsequent reduction in habitat heterogeneity may occur through direct (e.g. river sand extraction) or indirect (e.g. through surface runoff from logging) impacts (Dudgeon et al. 2006). For example, increasing sediment runoff may decrease plant diversity, remove organic debris, decrease habitat complexity, destroy breeding grounds, and remove shelter from

predators and habitat for prey species (Giam et al. 2015). The drainage of wetlands in SEA for oil palm monoculture degrades natural habitats and leads to a loss of ecosystem services. In particular, peat swamps in SEA formerly constituted 60% of known tropical peatlands (Posa et al. 2011), and their diverse and highly endemic fauna are threatened due to draining, agricultural development and logging activities (e.g. De Grave et al. 2015. Peat swamps have been reduced from 77% of their original cover in 1990 to only 36% cover in 2010, and may even be completely cleared by 2030 (Miettinen et al. 2012). As well as regulating water flow, stabilizing evaporation rates, and supporting endemic flora and fauna, these FEs act as significant carbon sinks, containing around nine times the carbon released globally by fossil-fuel combustion in 2006; their clearance would increase carbon emissions and fire risk would become greater as the swamps are drained (Miettinen et al. 2012). The 2015 forest fires in Indonesian Borneo were caused by draining peat swamps to create palm oil plantations in the context of a severe El Niño event. The fires caused the displacement and deaths of humans and animals, 500,000 cases of acute respiratory tract infections, and a loss of ~US $30 billion to the Indonesian economy (Lamb 2015), while affecting neighbouring SEA countries as well. Protection of FEs can therefore be seen as a self-interested insurance policy against climate change and natural disasters (Miettinen et al. 2012).

3.4 Over exploitation

Compared to other threat categories that affect all freshwater biodiversity from microbes to megafauna, over exploitation mostly affects vertebrates (fish, reptiles and some amphibians) (Dudgeon et al. 2006). Overexploitation of flora and fauna and subsequent reduction in biodiversity is linked to ecosystem disruption, increased disease risk, decreased fisheries yield and increased yield variability (Brooks et al. 2016). Fisheries are under pressure due to excessive demands for food, market pressures, enhanced fishing gear, weak or non-existent

management policies, and accidental by-catch, while the aquarium trade in wild species is now the most rapidly growing agricultural sector (Reid et al. 2013).

However, fish are not the only example of over exploitation. Deforestation rates in SEA are among the highest on Earth (1.2-2% per year (UNEP 2009)), mainly through logging, mining, and the creation of palm oil and rubber plantations and rice agriculture (Giam et al. 2015, Richards and Friess 2015), with mangrove forests also deforested at rates of 0.18% per year (Richards and Friess 2015). Intensive agriculture centred on crop monoculture such as oil-palm plantations presents one of the biggest threats to biodiversity in SEA, with knock-on effects on water pollution and habitat degradation as discussed above (UNEP 2009; Miettinen et al. 2011; Giam et al. 2015).

3.5 Species invasion

Invasive species present one of the most significant, inadequately controlled, and least reversible of threats to FEs, with impacts on ecosystems and their services such as eutrophication, reduction of biodiversity, alteration of fire regimes, destruction of fisheries and introduction of disease (Peh 2010; Allen et al. 2012). Invasive species in SEA are introduced from a range of sources including aquaculture (Water Hyacinth, Apple Snails, Tilapia), pest control (Mosquitofish) or the aquarium and pet trade (Armoured Catfish) (Peh 2010; Allen et al. 2012; Reid et al. 2013). Many examples exist in SEA where impacts of invasive species are observed at the physiochemical, trophic, and habitat level. For example, bioturbation and siltation is caused by Common Carp, community composition is altered by predation on fish by invasive Snakehead species, and habitat structure is impacted by Water Hyacinth and Floating Fern, which prevent movement of fishing boats and cause fishing net entanglement. Invasive species are more successful in degraded habitats (Allen et al. 2012),

and so their impact will likely be compounded as FEs in SEA are further degraded by human activities (see above) and the climate continues to warm (Peh 2010).

3.6 Global change

The above threats are affected by environmental changes occurring at the global scale such as nitrogen deposition, climate warming and altered precipitation patterns (Dudgeon et al. 2006). Environmental change will exacerbate habitat loss in SEA by altering the seasonal (monsoonal) patterns of precipitation by diminishing the Himalayan water sources for many of the major continental SEA rivers (Xu et al. 2009), or through further reduction of freshwater catchment area as sea-level rise continues (Woodruff 2010). The effect of rising greenhouse gases may be compounded in FEs. A meta-analysis of seasonal variations in CH4 emissions in wetlands, rice paddies and aquatic ecosystems showed a significant increase in methane production by microorganisms in freshwaters with increasing temperatures (Yvon-Durocher et al. 2014). Ecosystem services in the tropics will be adversely affected by climate change, where increased natural disasters and temperatures will cause an increase in vector and water-borne diseases, and a decrease in fishery and agricultural productivity.

4. Protecting FEs in Southeast Asia: future directions

Conservation initiatives which aim to find sustainable ways to meet human needs whilst protecting biodiversity and ecosystems do exist, such as those highlighted by the UN's International Decade for Action "Water for Life" 2005-2015 (UN 2015). Initiatives that involve SEA include the creation of the River Basin Committees in Lao PDR, Payment for Forest Environmental Services and a Biodiversity Conservation Action Plan in Vietnam, the River System Rehabilitation and the creation of a Central Non-Revenue Water Division in

the Philippines, as well as the Rewards for Watershed Services in Indonesia (UN 2015). These 'payment for ecosystem services' schemes involve downstream users paying those upstream to protect ecosystems and thereby maintain the provision of services, with the added-value effect that biodiversity is protected upstream.

In addition to these initiatives, we introduce below a series of seven conservation solutions that can be implemented individually or in tandem, according to the circumstances and threats specific to an individual drainage basin or water body

4.1 Seven solutions to conserve Southeast Asian freshwater ecosystems

Solution 1) Identification and implementation of Freshwater Protected Areas (FWPAs)

The creation of new FWPAs must be developed by determining patterns of species richness and endemicity within FEs and challenging whether they mirror documented patterns of the terrestrial hotspots, used as evidence to create terrestrial protected areas (Herbert et al. 2010). FWPAs not only conserve biodiversity but human water security downstream, with the Mekong and parts of the Indo-Malaysian Peninsula noted for their level of threat and potential impact on their large downstream human communities (Harrison et al. 2016). Although FWPAs already exist in Southeast Asia, such as five Ramsar sites in Malaysia, the creation of more FWPAs would be a beneficial, if obvious, solution. However, the methods by which suitable sites are identified by researchers and then considered for protection by governments and policy-makers need improvement. For example, of the 2,227 Ramsar Wetlands of International Importance worldwide (IUCN 2016), SEA has 49 whilst Europe has 1,067, which when scaled up to total area is equivalent to a ratio of roughly 1:10, indicating the disparity in attention to wetlands in the two regions. Of the 49 Ramsar sites within SEA, only 44% have a management plan: 36% have no plan, and 18% have plans in preparation (Ramsar 2014). The lack of management plans, and disparity in commitment

between countries of SEA to protect FEs indicates a need for more effective implementation of FWPAs.

Where management bodies exist to protect freshwaters, in many cases they do not act effectively. For example, the failure of the four nation, intergovernmental Mekong River Commission (MRC) to influence either the construction of mainstream dams underscores such ineffectiveness (Dudgeon 2011), although the MRC concerns could have had some influence on the Lao PDR Government's decision to undertake an additional review of the potential impacts of the Xayaburi Dam (Stone 2011).

It has been shown that terrestrial protected areas do not always provide sufficient protection for FEs (Hermoso et al. 2015), which is unsurprising due to their lack of consideration of protection for headwaters, catchments, or downstream areas (Dudgeon et al. 2006; Darwall et al. 2009). Terrestrial and FEs differ greatly in their evolution, ecology and function, and unlike terrestrial protected areas, most FWPAs cannot be effectively implemented by delineating a perimeter around a patch of land alone. Conservation and management of FEs must therefore consider all activities within the drainage basin due to FEs high level of connectivity, and so rather than drawing geographical boundaries around an area of land, it would be more beneficial to FWPAs to consider basin-level protection (Darwall et al. 2009). The Freshwater Ecosystems of the World initiative (Abell et al. 2008) provides a good starting point, albeit on a relatively coarse scale. Water engineering approaches such as environmental water allocations address this, by mimicking the complexity and natural variability of freshwater ecosystems which can underpin conservation strategies (Arthington 2012). However, the creation of basin-wide FWPAs in SEA where human populations are dense may be impractical in reality, and so the combination of smaller-scale FWPAs with other solutions such as those listed below would be most effective.

Solution 2) The creation of basin level management authorities:

Major freshwater ecoregions often span multiple transnational boundaries (Dudgeon et al. 2006; Abell et al. 2008), as does the Mekong River in SEA, and development proposals or conservation efforts at (or below) the national level often conflict with regional concerns or priorities (Chellaney 2011). Management of FEs cannot draw political boundaries based on countries when multiple countries often share watersheds (see Box 1), and so international water cooperation must be encouraged, utilising resources such as the recently developed Transboundary Waters Assessment Program (TWAP 2016). Institutions that permit and finance activities that affect FEs should utilise basin-scale analyses, and subsequently require basin-level management authorities. Such authorities would account for the joint-interests of multiple countries within the basin, cumulative impacts between one country and another, and the effects of universal pressures such as climate change (Winemiller et al. 2016). The role of political actors, social movements, implementing groups, power-brokers and consumers must be considered when assessing the political feasibility of a given ecological response strategy. In addition, multiple aspects of the project must be considered, including sustaining ecosystem services, biodiversity conservation, and human livelihoods (Winemiller et al. 2016; Welcomme et al. 2016). A basin level management authority would unite these multiple scales of decision-making. For example, the decision to build a dam to protect a flood plain in rural Cambodia should involve the people of the local adjacent village, the local and national government body coordinating the project, the financial organization funding the project, and downstream communities and countries affected by the dam. If stakeholders with shared interest in water resources and the ecosystem services provided by FEs do not cooperate effectively, whether on a village-to-village or country-to-country basis, conflict is bound to arise. As water scarcity is set to become Asia's defining crisis by mid-

century (Chellaney 2011), the need for inter-country cooperation across basins is more important than ever. The complex case of the Greater Mekong Basin is discussed in Box 1, a basin which could be united by the Mekong River Commission's basin-level management authority treaty. However, the lack of full participation by China and Myanmar remains problematic.

Solution 3) Protect human interests

By reaching a compromise between biodiversity conservation, ecosystem functioning and human livelihoods, freshwater conservation may be more successful in the long-term. The maintenance of healthy FEs is also possible through changing land use practises in ways that are also economically and socially beneficial such as the payment for ecosystem services schemes mentioned above. For example, although agriculture presents a threat to FEs, this threat can be alleviated by the creation of relatively narrow riparian reserves (Giam et al. 2015). In Kalimantan, Indonesia, forested riparian reserves maintained richness and functional diversity of stream fish communities whereas plantations with no riparian reserves had lower species richness and biomass. Oil palm growers should therefore be routinely mandated to create riparian reserves in oil-palm plantations, a measure that is supported by the Indonesian Sustainable Palm Oil system introduced in 2011 (Giam et al. 2015). Although these practises are by no means a 'get-out-clause' for habitat destruction, by incorporating integrative and sustainable approaches such as these more widely, the impact of agricultural threats to FEs within SEA may be reduced.

Solution 4) Enhancement of freshwater research capacity

Improved surveys and inventorying are urgently required to identify candidate sites for implementation of initial conservation policy for FWPA designation. A recent global

assessment of freshwater biogeographic regions or 'ecoregions' based on freshwater fishes and herpetofauna provides initial data for SEA, but the data quality is too poor for much of the region to make the required extrapolations (Abell et al. 2008).

By focusing research on keystone species which provide important ecosystem services, attention is more likely to be attracted to conservation. For example, freshwater mussels such as *Lamellidens marginalis* (found in Northern SEA), act as microhabitat engineers, performing crucial ecological functions such as filtering water, transporting nutrients and oxygenating sediments, allowing biodiversity to thrive even where pollutant levels are high (Chowdhury et al. 2016). Flagship species such as the Irrawaddy river dolphin, or the Mekong giant catfish could be used to enhance public consciousness of freshwater fishes as a starting point for more wide-ranging conservation initiatives combining efforts of scientists and natural resource managers (Welcomme et al. 2016).

Spatial variation in data availability within biodiversity databases has been explained by low per capita gross domestic product (GDP), low level of English speakers, geographical distance away from the country hosting the database holding the information of interest and degree of civil or international conflict (Amano and Sutherland 2013). This is relevant to SEA, where GDP is low, databases have poor records (such as the Global Biodiversity Information Facility which has 0.004 – 0.6 records/km^2 in SEA compared with 252/km^2 in the United Kingdom). In addition, although local scientists are able to work in their own language, English literacy is poor (Singapore being the exception), and there are problems with civil and international conflict (Chellaney 2011; Amano and Sutherland 2013; Schatz 2014).

Solution 5) Fix the leak

Human water security in SEA (Figure 2) depends on adequate water supplies; the more water that is wasted, the more must be appropriated from nature to meet these needs. Up to 50% of water can be lost from the distribution system of cities due to leaks and other problems: Taiwan, for example, loses almost 2 million cubic meters per day due to leakage. If leaks such as these were prevented, water waste reduced, and water conservation encouraged, governments and local authorities would not need to invest additional finances creating dams and treatment plants. In addition, water 'productivity' can be increased by 1) replacing higher-quality water with lower-quality water; 2) purifying lower-quality water into higher-quality water by treatment; and 3) decreasing the volume of higher-quality water used to create goods and services (Grant et al. 2012). This would decrease destruction or degradation of FEs (and impacts on biodiversity, Figure 5) by reducing water abstraction from rivers and lakes and the need for irrigation dams. The combination of these approaches would lessen the pressure on the exploitation and subsequent degradation of FEs.

Solution 6) Respect and utilise religious, political and consumer power

Different parts of SEA have a variety of political regimes, religions and societal behaviours, illustrated by Buddhist Thailand, Muslim Indonesia and more secular Vietnam. If scientists are to work towards FE conservation in SEA, they must carefully consider the political, religious and societal factors that are likely to vary vastly between countries. In SEA, religious motivation may provide a more effective basis for conservation of protected areas than government edict, with some religions and sacred spaces having a deep connection to the natural world (Taylor 2012). Some sacred natural sites have higher levels of biodiversity than surrounding areas, due to their remnant nature and high level of protection, and may even harbor species that are extinct in the wild (Taylor 2012).

The Indonesian Council of Ulama issued a fatwa (an Islamic call to action) in 2014 for Muslims to take an active role in protecting and conserving the endangered species of Indonesia (Actman 2015). This fatwa, which was the first of its kind, was then joined by the state of Terengganu in Malaysia in 2015 in condemnation of wildlife poaching. Collaboration between scientists and Islamic leaders for the conservation of FEs holds great potential in SEA, where around 235 million people — constituting 12.7% of the world's Muslims in Indonesia alone (Pew Research Centre, 2016) — could be engaged.

Utilising governmental and consumer influence could also benefit FE conservation. Applying pressure to hold companies and governments accountable for environmental damage has become increasingly possible in the digital era, such as the online community purchase of 389 acres of Bornean rainforest by the petition website 'Avaaz', helping to protect 700 of Borneo's remaining orang-utans, and 300 pygmy elephants. Indeed, companies and governments who invest in such solutions could be at a competitive advantage when targeting modern consumers who are prepared to pay for environmentally friendly products, which could be created using a similar approach to sustainability credits, but with FEs in mind. When consumers direct spending to companies who invest in such initiatives, and campaign against companies who damage the environment, such as campaigning for sustainable oil palm in food products, direct environmental protection may be achieved. If such campaigns could also be used to protect iconic freshwater species such as the Irrawaddy dolphin in the Mahakam River, Borneo, this would provide another opportunity for preserving FEs.

Pressure from local communities within SEA as well as the global online community can create a force for change, such as that driven by environmental campaigner Jintana Kaewkao and fellow villagers who blocked the construction of a major coal-fired power plant in Ban Krut, Thailand, and are now campaigning to block plans for a steelwork plant on a

wetland (Schatz 2014). However, environmental campaigning in some SEA countries can be dangerous: 16 Thai environmentalists were murdered between 2002 and 2013 (Schatz 2014), illustrating the risks facing by conservation activists in the region.

Solution 7) Utilise new technologies

Consumer, religious and political power, as discussed above, can be enhanced by the immediacy of online software, social media and smartphone devices. Free software such as the Spatial Monitoring and Reporting Tool, Cybertracker and the Biodiversity Indicators Dashboard by Nature Serve (Bhammar 2014) can be used by everyone from local communities to governments, for everything from scientific research to crime prevention, poaching activity, and monitoring of butterfly distributions, for example. These programs allow holistic, integrative GPS field data documentation and visualization on key biodiversity indicators, enabling the tracking of biodiversity and conservation performance to help track progress toward conservation targets, national monitoring, outcome-based policy making and catalyse necessary investments in information infrastructure. New online databases such as Fishes of Mainland Southeast Asia (FiMSEA) (Kano et al. 2013) allow open access data to be collected, shared and analysed. New modelling frameworks such as GLOBIO-Aquatic (Janse et al. 2015) allow users to assess impacts of human induced environmental drivers, or predict trends under future scenarios using spatial information on environmental drivers and cause-effect relationships derived from literature. The increasing availability of spatial data on biodiversity and ecosystem services supports sophisticated trade-off analyses which can inform assessment protocols for developments such as hydropower dams (Winemiller et al. 2016), avoiding potential negative impacts of such projects.

Developments in molecular methods also have potential for FE conservation, such as the use of environmental DNA, which has the potential to offer an efficient, reliable and

informative method for monitoring FEs of SEA, as it is particularly suited to detecting elusive, endangered or invasive species in aquatic environments (Bohmann et al. 2014). By utilising new technologies such as these, more information can be rapidly gained and shared with the communities and policy-makers who influence the protection of FEs of SEA.

5. Conclusions

We outline above seven key solutions to the ongoing degradation of SEA freshwater ecosystems. The most critical of these, in the short term, is the creation of basin level management authorities. Unless regional basin management authorities can unite to face the problems discussed above, we foresee a permanent and irreversible loss of biotic, environmental, societal, and economic assets.

Figure Legends

Figure 1: Global species richness maps for freshwater species from (Collen et al. 2013). Upper map shows the total normalised species richness and the lower map shows the normalised species richness of threatened species.

Figure 2: Map showing the adjusted human water security (HWS) incident threat in Southeast Asia (constructed using http://www.riverthreat.net, see Vörösmarty et al. 2010, Figure 1).

Figure 3: Indicators of freshwater biodiversity richness across four Southeast Asian biodiversity hotspots. Panel charts of richness and endemism of five freshwater groups (fishes, amphibians, crabs, turtles, crocodiles) within four biodiversity hotspots (Indo-Burma, Philippines, Sundaland, Wallacea). Upper panels (small scale view) show complete bars for all groups; lower panels (large scale view) show close-up of base of bar for fish and complete bars for other groups. Red and blue colours indicate endemic and non-endemic species, respectively. Figures above each bar represent total species richness, with percentage endemism in parentheses. Data sources: Buhlmann et al. (2009), Freshwater Ecoregions of the World (http://www.feow.org/), and AmphibiaWeb (http://amphibiaweb.org/), and unpublished data (D.J. Yeo).

Figure 4: Fish diversity and dam count in the Mekong (from Winemiller et al. 2016). White dots illustrate dams under construction or already built. Red dots illustrate dams planned by 2030.

Figure 5: Combined threats to biodiversity in SEA arising from pollution, drainage-basin alteration, flow regulation, overexploitation and exotic fishes (constructed using http://www.riverthreat.net, see Vörösmarty et al. 2010).

References:

1. Abell R, et al. 2008. Freshwater ecoregions of the world: a new map of biogeographic units for freshwater biodiversity conservation. BioScience 58: 403–414.
2. Actman J. 2015. Muslim Council Issues Fatwa Against Poaching. Widlife Watch. National Geographic. (16 December 2015; http://news.nationalgeographic.com/2015/12/151216-fatwa-terengganu-malaysia-poaching/)
3. Allen DJ, Smith KG, and Darwall WRT. 2012. The status and distribution of freshwater biodiversity in Indo-Burma. IUCN.
4. Amano T, and Sutherland WJ. 2013. Four barriers to the global understanding of conservation: wealth, language, geographical location and security. Proceedings of the Royal Society B: Vol. 280. No. 1756.
5. Arthington, A.H. (2012) Environmental Flows: Saving Rivers in the Third Millennium. University of California Press, Oakland.
6. Balian EV, Segers H, Lévèque C and Martens K. 2008. The Freshwater Animal Diversity Assessment: an overview of the results. Hydrobiologia 595:627–637
7. Bhammar H. 2014. 20 NEW AND POWERFUL CONSERVATION TOOLS. F&ES BLOG. Yale school of forestry and environmental sciences. (6 December 2014; http://environment.yale.edu/blog/2014/12/20-new-and-powerful-conservation-tools/)
8. Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, Yu D, and de Bruyn, M. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. Trends in Ecology & Evolution 29(6): 358–367.
9. Brooks EGE, Holland RA, Darwall WRT, and Eigenbrod F. 2016. Global evidence of positive impacts of freshwater biodiversity on fishery yields. Global Ecology and Biogeography 25: 553-562.

10. Buhlmann KA, Akre TSB, Iverson JB, Karapatakis D, Mittermeier RA, Georges A, Rhodin AGJ, van Dijk PP and Gibbons JW. 2009. A global analysis of tortoise and freshwater turtle distributions with identification of priority conservation areas. Chelonian Conservation and Biology 8: 116–149.
11. Chellaney B. 2011. Water: Asia's new battleground. Georgetown University Press.
12. Chowdhury GW, Zieritz A and Aldridge DC. 2016. Ecosystem engineering by mussels supports biodiversity and water clarity in a heavily polluted lake in Dhaka, Bangladesh. Freshwater Science 35: 188–199.
13. Cochard R. 2017. Coastal Water Pollution and Its Potential Mitigation by Vegetated Wetlands: An Overview of Issues in Southeast Asia. Redifining Diversity and Dynamics of Natural Resource Management in Asia, Volume 1. Elsevier Inc.
14. Collen B, Whitton F, Dyer EE, Baillie JE, Cumberlidge N, Darwall WRT, Pollock C, Richman NI, Soulsby AM, Böhm M. 2014. Global patterns of freshwater species diversity, threat and endemism. Global Ecology and Biogeography 23(1): 40–51.
15. Darwall W, Smith K, Allen D, Seddon M, McGregor Reid G, Clausnitzer V, and Kalkman VJ. 2009. Freshwater biodiversity – a hidden resource under threat. Wildlife in a changing world: an analysis of the 2008 IUCN Red List of Threatened Species: 43–54.
16. De Grave et al. 2015. Dead shrimp blues: A global assessment of extinction risk in freshwater shrimps (Crustacea: Decapoda: Caridea). Public Library of Science One 10(3): e0120198.
17. Dudgeon D et al. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews 81: 163–182.

18. Dudgeon D. 2011. Asian river fishes in the Anthropocene: threats and conservation challenges in an era of rapid environmental change. Journal of Fish Biology 79: 1487–1524.

19. Dudgeon D. 2014. Accept no substitute: biodiversity matters. Aquatic Conservation: Marine and Freshwater Ecosystems 24(4): 435–440.

20. Dugan PJ et al. 2010. Fish Migration, Dams, and Loss of Ecosystem Services in the Mekong Basin. Ambio 39(4): 344–348.

21. Giam X, Hadiaty RK, Tan HH, Parenti LR, Wowor D, Sauri S, Chong KY, Yeo DCJ and Wilcove DS. 2015. Mitigating the impact of oil-palm monoculture on freshwater fishes in Southeast Asia. Conservation Biology 29: 1357–1367.

22. Grant SB et al. 2012. Taking the "waste" out of "wastewater" for human water security and ecosystem sustainability. Science 337: 681–686.

23. Greer B, Maul R, Campbell K and Elliott CT. 2017. Detection of freshwater cyanotoxins and measurement of masked microcystins in tilapia from Southeast Asian aquaculture farms. Analytical and Bioanalytical Chemistry: pp.1-13.

24. Hails C et al. 2008. Living Planet Report 2008. WWF.

25. Harrison IJ, Green PA, Farrell TA, Juffe-Bignoli D, Sáenz L, and Vörösmarty CJ. 2016. Protected areas and freshwater provisioning: a global assessment of freshwater provision, threats and management strategies to support human water security. Aquatic Conservation: Marine and Freshwater Ecosystems: 26(S1) 103-120.

26. Herbert ME, McIntyre PB, Doran PJ, Allan JD and Abell R. 2010. Terrestrial reserve networks do not adequately represent aquatic ecosystems. Conservation Biology 24: 1002–1011.

27. IUCN. 2016. Freshwater. IUCN WCPA Freshwater Specialist Group (3 March 2017; https://www.iucn.org/theme/protected-areas/wcpa/what-we-do/freshwater)

28. Janse JH, Kuiper JJ, Weijters MJ, Westerbeek EP, Jeuken MHJL, Bakkenes M, Alkemade R, Mooij WM. and Verhoeven JTA. 2015. GLOBIO-Aquatic, a global model of human impact on the biodiversity of inland aquatic ecosystems. Environmental Science & Policy 48: 99-114.

29. Kano Y, Adnan MS, Grudpan C, Grudpan J, Magtoon W, Musikasinthorn P, Natori Y, Ottomanski S, Praxaysonbath B, Phongsa K and Rangsiruji A. 2013. An online database on freshwater fish diversity and distribution in Mainland Southeast Asia. Ichthyological Research 60(3): p.293.

30. Kano Y, Dudgeon D, Nam S, Samejima H, Watanabe K, Grudpan C, et al. (2016) Impacts of Dams and Global Warming on Fish Biodiversity in the Indo-Burma Hotspot. PLoS ONE 11(8): e0160151.

31. Kareiva P and Marvier M. 2012. What is conservation science? BioScience 62 (11): 962–969.

32. Lamb K. 2015. Indonesia's fires labelled a 'crime against humanity' as 500,000 suffer. Indonesia. World. The Guardian. (26 October 2015; http://www.theguardian.com/world/2015/oct/26/indonesias-fires-crime-against-humanity-hundreds-of-thousands-suffer

33. Lévêque C, Oberdorff T, Paugy D, Stiassny MLJ and Tedesco PA. 2008. Global diversity of fish (Pisces) in freshwater. Hydrobiologia 595(1): 545–567.

34. Miettinen J, Shi C and Liew SC. 2012. Two decades of destruction in Southeast Asia's peat swamp forests. Frontiers in Ecology and the Environment 10(3): 124–128.

35. Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB, Kent J. 2000. Biodiversity hotspots for conservation priorities. Nature 403: 853–858.

36. Peh KSH. 2010. Invasive species in Southeast Asia: the knowledge so far. Biodiversity and Conservation 19: 1083–1099.

37. Pew Research Centre. 2016. Table: Muslim Population by Country. Religion & Public Life. Pew Research Centre. (27 January 2011; http://www.pewforum.org/2011/01/27/table-muslim-population-by-country/#)

38. Poff NL and Zimmerman JK. 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. Freshwater Biology 55(1): 194–205.

39. Posa MRC, Wijedasa LS and Corlett RT. 2011. Biodiversity and conservation of tropical peat swamp forests. BioScience 61(1): pp.49-57.

40. Ramsar. 2014. About The Ramsar Convention. Retrieved at http://www.ramsar.org/about-the-ramsar-convention.

41. Reid GMcG, Contreras MacBeath T and Csatadi K. 2013. Global challenges in freshwater fish conservation related to public aquariums and the aquarium industry. *International Zoo Yearbook* 47(1): 6–45.

42. Rowley J, et al. 2010. Impending conservation crisis for Southeast Asian amphibians. Biology Letters 6: 336–338.

43. Schatz, JJ. 2014. 2015. Thai environmentalists pay for activism with their lives. Aljazeera (5 April 2014; http://projects.aljazeera.com/2015/04/thailand-activists/)

44. Scott D. 1989. A directory of Asian Wetlands. IUCN, Cambridge.

45. Stone R. 2011. Mayhem on the Mekong. Science 333: 814.

46. Taylor KAM. 2012. Sacred natural sites. Earthscan.

47. Triet T, et al. 2014. Persistent Organic Pollutants in Wetlands of the Mekong Basin. U.S Geological Survey. Report 2013–5196.

48. TWAP (Transboundary Waters Assessment Programme). 2016. River basins Component. (19 January 2016; http://twap-rivers.org/)

49. UN 2015. "UN Water and Sanitation Best Practices Platform". International Decade for Action "Water for Life" 2005–2015. (http://www.unwaterbestpractices.org/)

50. UNEP. 2009. Vital Forest Statistics. United Nations Environment Programme (UNEP), United Nations Food and Agriculture Organisation(FAO) and United Nations Forum on Forests (UNFF). UNEP Job No: DEW/1032/NA.

51. Vörösmarty CJ, et al. 2010. Global threats to human water security and river biodiversity. Nature 467: 555–561.

52. Welcomme, R.L., Baird, I.G., Dudgeon, D., Halls, A., Lamberts, D. and Mustafa, M.G., 2016. Fisheries of the rivers of Southeast Asia. Freshwater Fisheries Ecology, pp.363-376.

53. Winemiller KO, et al. 2016. Balancing hydropower and biodiversity in the Amazon, Congo, and Mekong. Science 351(6269): 128–129.

54. Woodruff D. 2010. Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. Biodiversity and Conservation 19: 919–941.

55. Worldometers. 2017. "South-eastern Asia Population". (http://www.worldometers.info/world-population/south-eastern-asia-population/)

56. Xu J, Grumbine RE, Shrestha A, Eriksson M, Yang X, Wang Y, Wilkes A. 2009. The melting Himalayas: cascading effects of climate change on water, biodiversity, and livelihoods. Conservation Biology 23: 520–530.

57. Yvon-Durocher G, Allen AP, Bastviken , Conrad R, Gudasz C, St-Pierre A, Thanh-Duc N and Del Giorgio PA. 2014. Methane fluxes show consistent temperature dependence across microbial to ecosystem scales. Nature *507* (7493); 488–491.

58. Ziv G, Baran E, Nam S, Rodriguez-Iturbe I, Levin SA. 2012. Trading-off fish biodiversity, food security, and hydropower in the Mekong River Basin. Proceedings of the National Academy of Sciences USA 109: 5609–5614.

Acknowledgments: We thank Darren C. J. Yeo for construction of Figure 3.

Box 1: International conflict and environmental status of the Mekong Basin

River: The Mekong (= Lancang Jiang in China).

Length*:* 4,800 km.

Catchment area*:* 795,000 km^2

Countries: China, Myanmar, Cambodia, Laos, Thailand and Vietnam.

Species: > 2,200 new species since 1997. > 430 mammals, ~ 1,200 birds, > 800 reptiles & amphibians , > 1,100 fish (including four of the world's top 10 largest freshwater fish, > 20,000 plants.

Population**:** ~ 300 million people. ~ 80% depend on the ecosystem for food security, livelihoods and culture. Livelihoods of ~2.5 million people and 25% of Cambodia's protein from the Tonlé Sap alone.

Special features: Longest river in SEA. World's highest-yielding inland fishery with 2nd-3rd highest fish richness globally. Greatest extent of combined tiger habitat on Earth. Tonlé Sap (which has an associated Ramsar site) is the largest natural lake in SEA.

Pollution: Serious problem with endosulfan and its metabolites, POPs found in hotspot sites and important wetlands with concentrations exceeding ecological risk thresholds. Tonlé Sap contains DDT and DDE exceeding the Canadian and U.S. standards, with fauna also exhibiting high degrees of bioaccumulation.

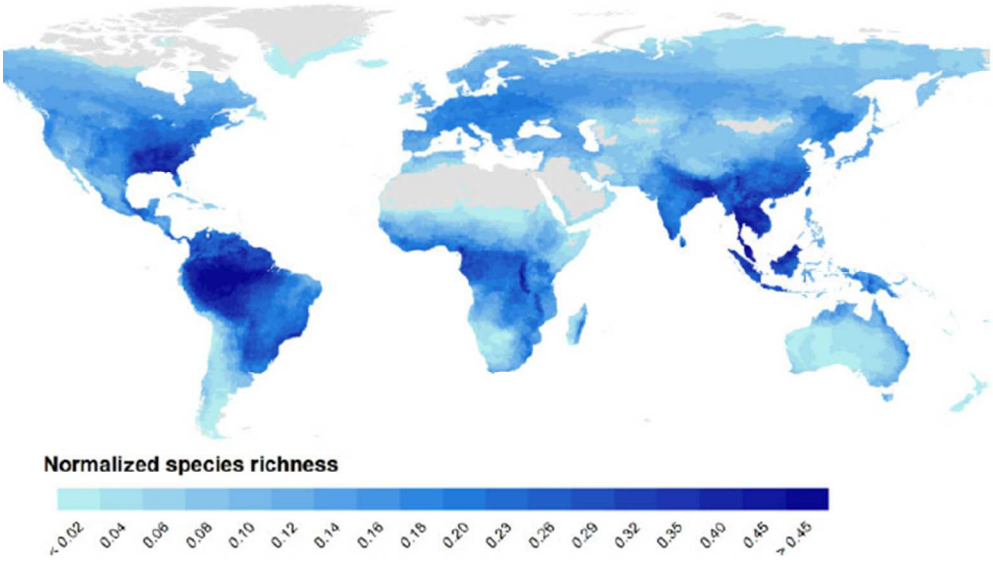
Climate change: Particularly vulnerable to climate change. Lower flooding levels, sea-level rise and hotter temperatures cause saline intrusion of the Mekong River, creating agricultural damage and a loss of fish species richness and abundance, impacting fisheries.

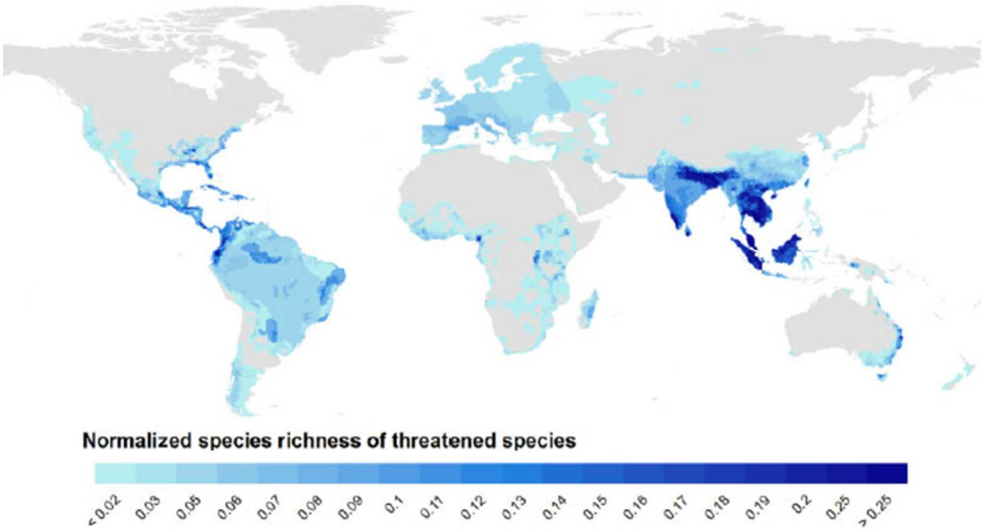
Habitat destruction: Greatest planned rate of growth of hydropower in the world: 98 proposed dams expected to damage riparian communities and aquatic biota. 70% of fish migration expected to be blocked by 11 dams from Laos to Cambodia alone in a region where the majority of the economy and livelihoods rely on fishing. Dams planned for the Lancang region could trap ~50% of sediment coming from China, blocking nutrient flow and changing river hydrology downstream.

Human conflict: Threats to the Mekong have potential to affect communities, flood dependent agriculture and river biota, possibly creating millions of environmental refugees. Conflicting political regimes and socio-economic needs between countries, combined with a North-South 'asymmetry of power' between authoritative China in the north and vulnerable developing countries to the south could create conflict. Laos, the largest contributor to the Mekong's flow, prioritises hydroelectric power; Vietnam, SEAs major rice producer, prioritises irrigation; Cambodia, with the sensitive Tonlé Sap depending on annual flood pulse cycles, prioritises conservation of the Mekong's unique hydrology; Thailand has multiple priorities of energy, irrigation and fisheries, whilst Myanmar's main relationship to

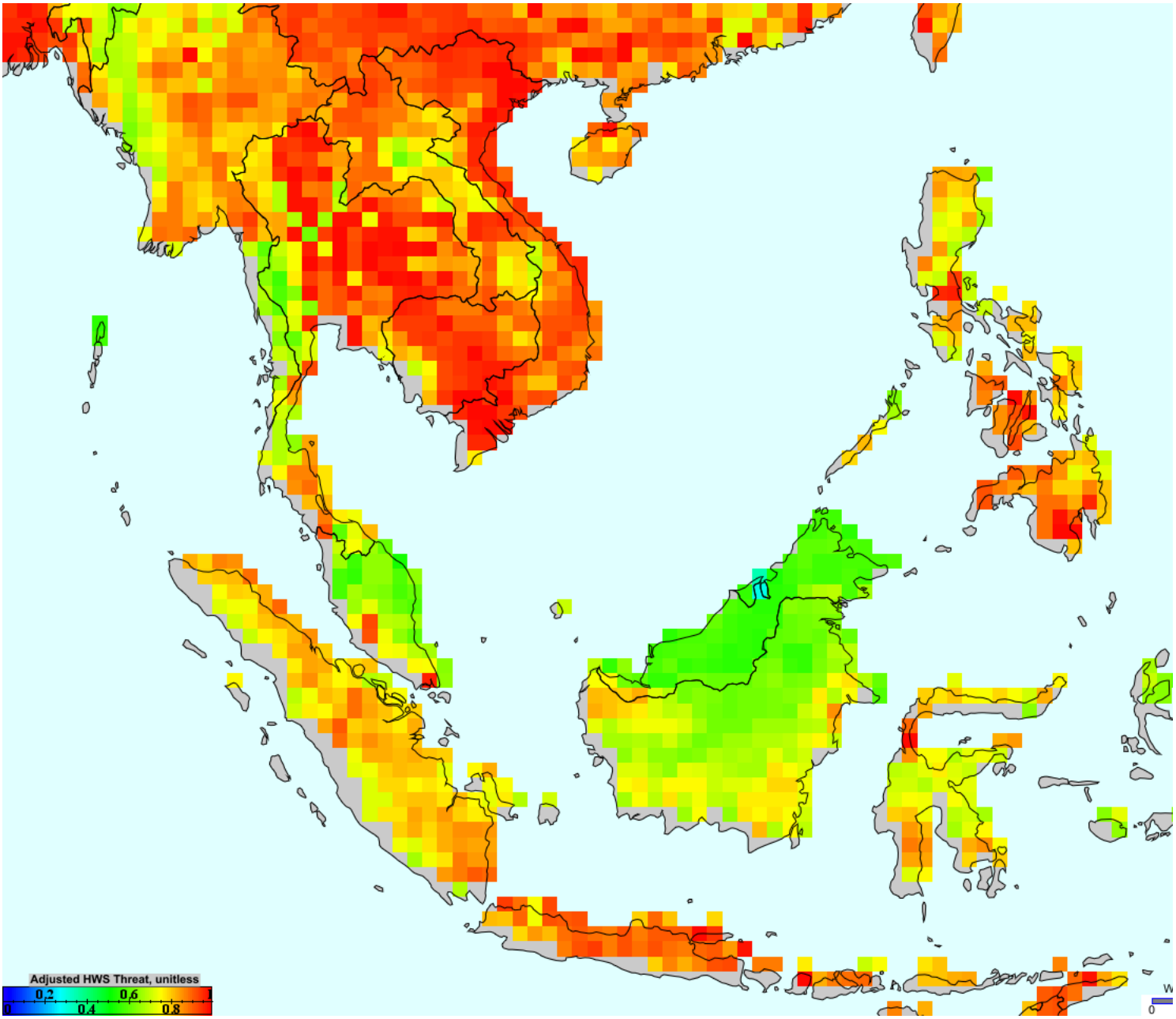
the Mekong is its political border. China's continued upstream activities and lack of willingness to engage more widely on their potential impacts may spark international conflict in the Mekong Basin.

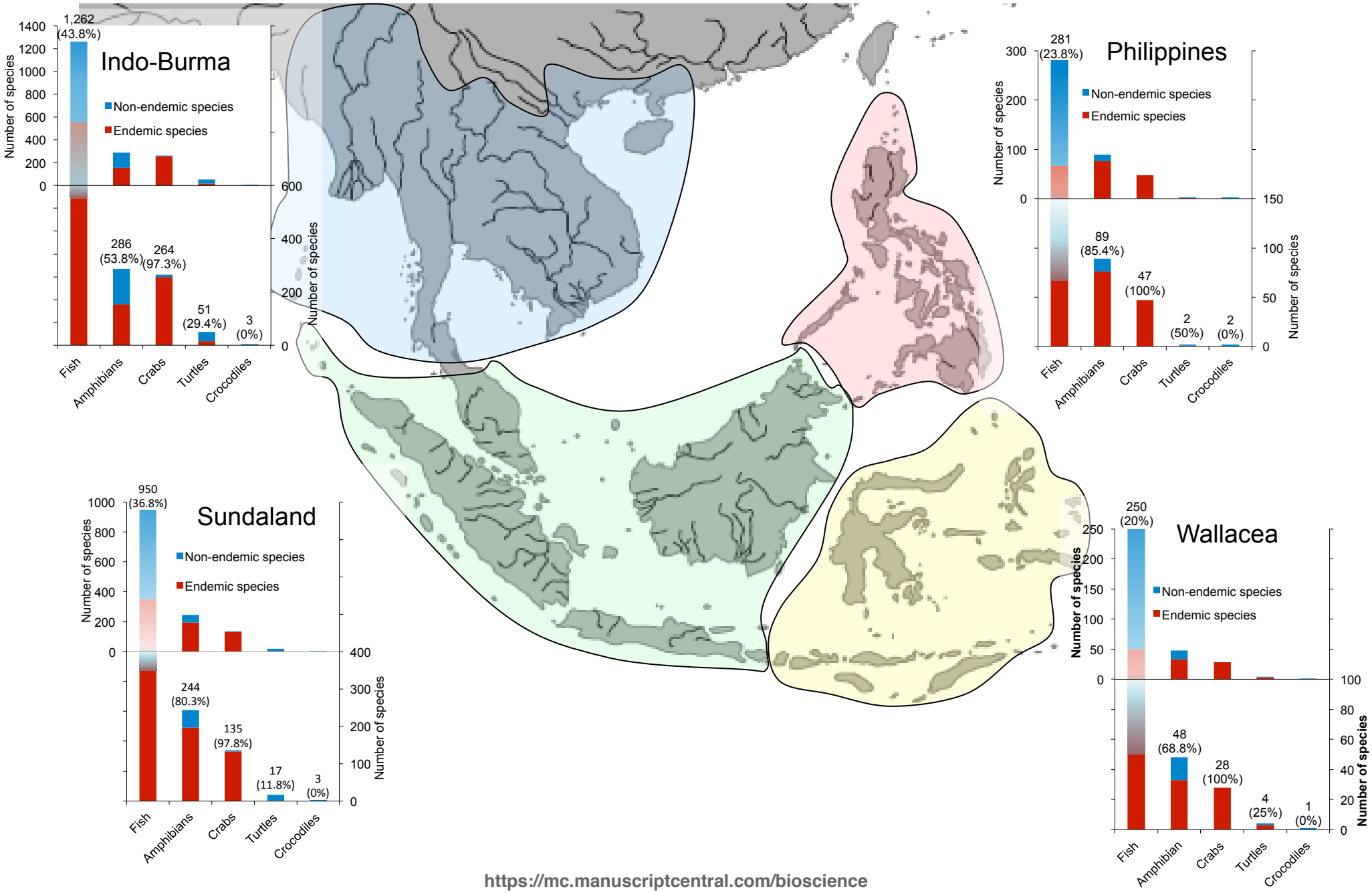
(Data sources: Triet et al. 2014; Chellaney, 2011; Stone 2011; Ziv et al. 2012)

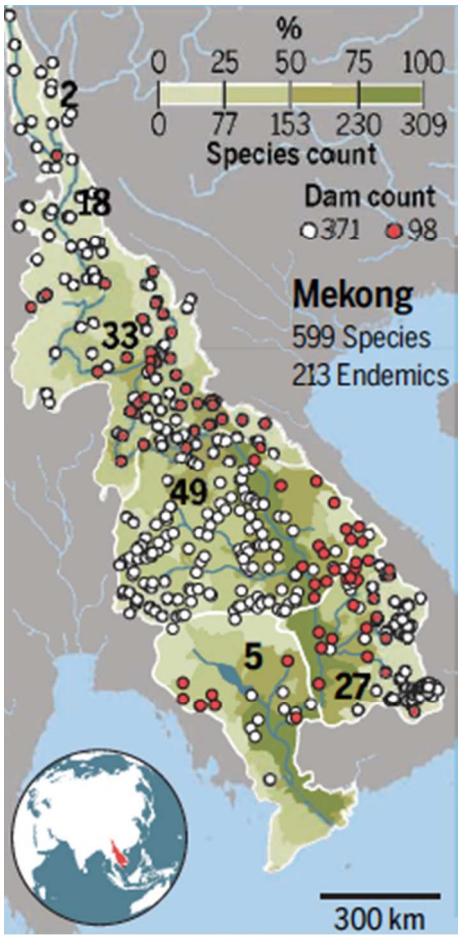




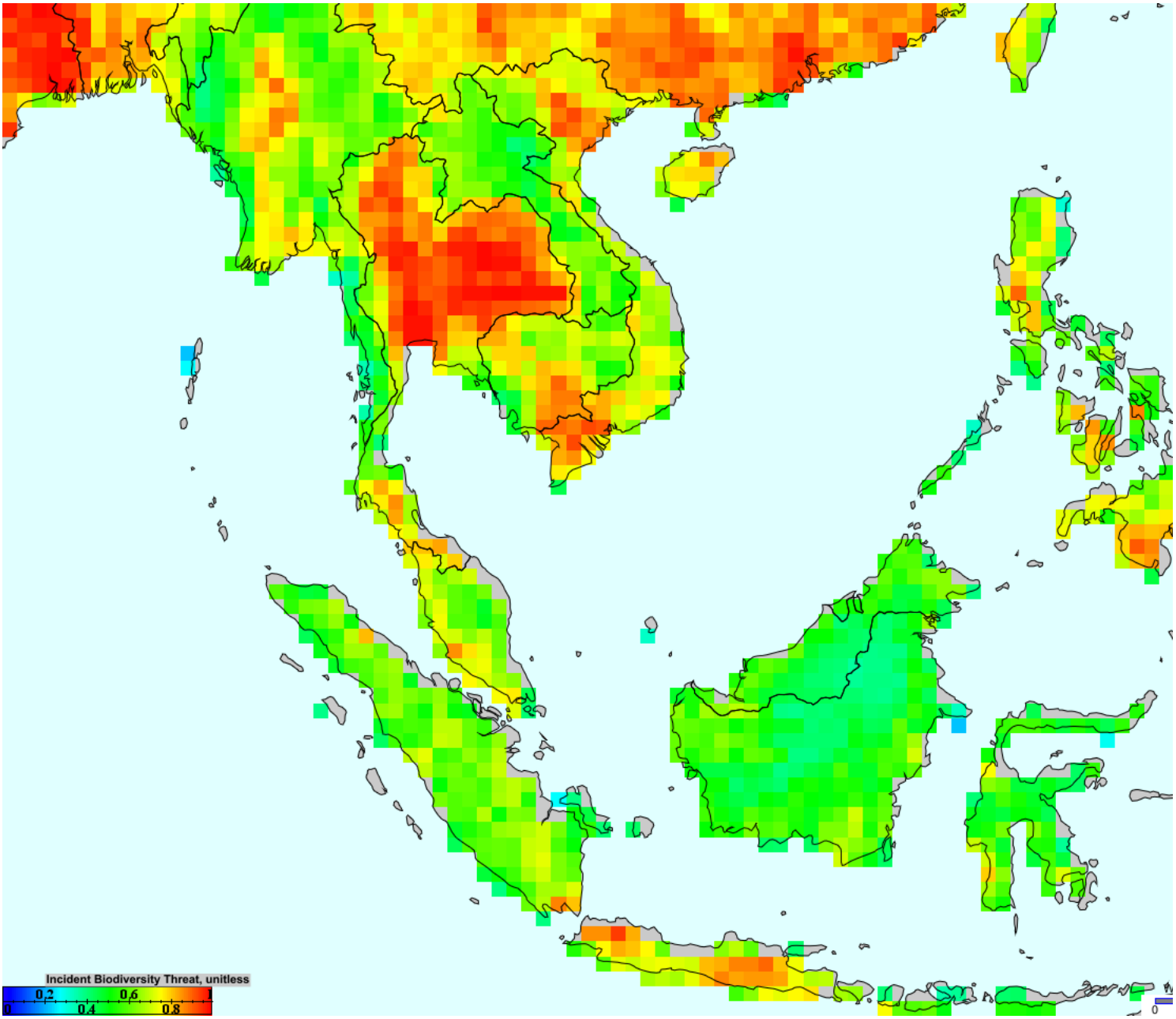
233x282mm (72 x 72 DPI)


Adjusted HWS Threat, unitless
0
0.2
0.4
0.6
0.8
1


Indo-Burma
Non-endemic species
Endemic species
Number of species
1,262 (43.8%)
286 (53.8%)
264 (97.3%)
51 (29.4%)
3 (0%)
Fish
Amphibians
Crabs
Turtles
Crocodiles
Philippines
281 (23.8%)
89 (85.4%)
47 (100%)
2 (50%)
2 (0%)
Sundaland
950 (36.8%)
244 (80.3%)
135 (97.8%)
17 (11.8%)
3 (0%)
Wallacea
250 (20%)
48 (68.8%)
28 (100%)
4 (25%)
1 (0%)
Amphibian



79x163mm (72 x 72 DPI)


Incident Biodiversity Threat, unitless
0
0.2
0.4
0.6
0.8
1

Appendix 4

Fish species known from study lakes

Species found in target lakes

Country	Region	Lake	Species	Reference
Indonesia	Bali	Batur	*Amphilophus sp. 'Black louhan'*	Sentosa and Wijaya, 2012, Budiasa *et al* . 2018
Indonesia	Bali	Batur	*Amphilophus sp. 'Red louhan'*	Sentosa and Wijaya, 2012, Budiasa *et al* . 2018
Indonesia	Bali	Batur	*Barbodes binotatus*	Sentosa and Wijaya, 2012, Budiasa *et al* . 2018
Indonesia	Bali	Batur	*Barbonymus gonionotus*	Green *et al* . 1978
Indonesia	Bali	Batur	*Barbodes microps*	Green *et al* . 1978
Indonesia	Bali	Batur	*Channa striata*	Green *et al* . 1978
Indonesia	Bali	Batur	*Chanos chanos*	Sentosa and Wijaya, 2012, Budiasa *et al* . 2019
Indonesia	Bali	Batur	*Clarias batrachus*	Green *et al* . 1978
Indonesia	Bali	Batur	*Cyprinus carpio*	Green *et al* . 1978
Indonesia	Bali	Batur	*Ctenopharyngodon idella*	Kartamihardja, 2012
Indonesia	Bali	Batur	*Monopterus albus*	Sentosa and Wijaya, 2012, Budiasa *et al* . 2018
Indonesia	Bali	Batur	*Oreochromis mosambicus*	Sentosa and Wijaya, 2012; Green *et al* . 1978, Budiasa *et al* . 2018
Indonesia	Bali	Batur	*Oreochromis niloticus*	Sentosa and Wijaya, 2012; Suryaningtyas and Ulinuha, 2016, Budiasa *et al* . 2018
Indonesia	Bali	Batur	*Oreochromis sp.*	Sentosa and Wijaya, 2012
Indonesia	Bali	Batur	*Poecilia reticulata*	Sentosa and Wijaya, 2012; Green *et al* . 1978, Budiasa *et al* . 2018
Indonesia	Bali	Batur	*Barbodes lateristriata*	Sentosa and Wijaya, 2012, Budiasa *et al* . 2018
Indonesia	Bali	Batur	*Rasbora sp.*	Sentosa and Wijaya, 2012, Budiasa *et al* . 2018
Indonesia	Bali	Batur	*Xiphophorus helleri*	Sentosa and Wijaya, 2012, Budiasa *et al* . 2018
Indonesia	Bali	Batur	*Xiphophorus maculatus*	Green *et al* . 1978
Indonesia	Bali	Beratan	*Amatitlania nigrofasciata*	Sentosa *et al* . 2013; Whitten *et al* . 1996
Indonesia	Bali	Beratan	*Amphilophus citrinellus*	Sentosa *et al* . 2013
Indonesia	Bali	Beratan	*Anabas testudineus*	Sentosa *et al* . 2013
Indonesia	Bali	Beratan	*Barbonymus gonionotus*	Sentosa *et al* . 2013
Indonesia	Bali	Beratan	*Channa striata*	Whitten *et al.* 1996
Indonesia	Bali	Beratan	*Clarias batrachus*	Whitten *et al.* 1996; Green *et al.* 1978
Indonesia	Bali	Beratan	*Clarias gariepinus*	Whitten *et al* . 1996
Indonesia	Bali	Beratan	*Clarius sp.*	Sentosa *et al* . 2013
Indonesia	Bali	Beratan	*Colossoma macropomum*	Sentosa *et al* . 2013
Indonesia	Bali	Beratan	*Ctenopharyngodon idella*	Sentosa *et al* . 2013; Whitten *et al* . 1996

Indonesia	Bali	Beratan	*Cyprinus carpio*	Sentosa *et al* . 2013; Whitten *et al* . 1996
Indonesia	Bali	Beratan	*Hypostomus sp.*	Sentosa *et al* . 2013
Indonesia	Bali	Beratan	*Monopterus albus*	Green *et al* . 1978
Indonesia	Bali	Beratan	*Oreochromis mossambicus*	Whitten *et al.* 1996
Indonesia	Bali	Beratan	*Oreochromis niloticus*	Sentosa *et al* . 2013; Green *et al* . 1978
Indonesia	Bali	Beratan	*Osphronemus gouramy*	Sentosa *et al* . 2013
Indonesia	Bali	Beratan	*Osteochilus vittatus*	Sentosa *et al* . 2013; Whitten *et al* . 1996
Indonesia	Bali	Beratan	*Poecilia reticulata*	Sentosa *et al* . 2013; Whitten *et al* . 1996; Green *et al.* 1978
Indonesia	Bali	Beratan	*Poecilia sp.*	Whitten *et al.* 1996
Indonesia	Bali	Beratan	*Barbodes binotatus*	Sentosa *et al* . 2013; Whitten *et al* . 1996
Indonesia	Bali	Beratan	*Rasbora argyrotaenia*	Sentosa *et al* . 2013
Indonesia	Bali	Beratan	*Rasbora baliensis*	Whitten *et al.* 1996;
Indonesia	Bali	Beratan	*Barbodes lateristriata*	Rahman *et al* ., 2012; Sentosa et al. 2013; Whitten et al. 1996
Indonesia	Bali	Beratan	*Xiphophorus hellerii*	Sentosa *et al* . 2013
Indonesia	Bali	Beratan	*Xiphophorus maculatus*	Green *et al.* 1978
Indonesia	Bali	Buyan	*Amatitlania nigrofasciata*	Dahruddin *et al.* 2016, Restu *et al.* 2016
Indonesia	Bali	Buyan	*Anabas sp.*	Green *et al.* 1978
Indonesia	Bali	Buyan	*Barbonymus gonionotus*	Green *et al.* 1978
Indonesia	Bali	Buyan	*Barbodes microps*	Green *et al.* 1978
Indonesia	Bali	Buyan	*Channa striata*	Green *et al.* 1978
Indonesia	Bali	Buyan	*Clarias batrachus*	Green *et al.* 1978
Indonesia	Bali	Buyan	*Cyprinus carpio*	Restu *et al.* 2016; Green *et al.* 1978
Indonesia	Bali	Buyan	*Gambusia affinis*	Dahruddin *et al* . 2016
Indonesia	Bali	Buyan	*Helostoma sp.*	Green *et al.* 1978
Indonesia	Bali	Buyan	*Monopterus albus*	Green *et al.* 1978
Indonesia	Bali	Buyan	*Oreochromis mosambicus*	Restu *et al.* 2016; Green *et al.* 1978
Indonesia	Bali	Buyan	*Oreochromis niloticus*	Restu *et al.* 2016
Indonesia	Bali	Buyan	*Osteocillus vittatus*	Restu *et al.* 2016; Green *et al.* 1978
Indonesia	Bali	Buyan	*Poecilia reticulata*	Green *et al* . 1978
Indonesia	Bali	Buyan	*Rasbora sp.*	Green *et al* . 1978
Indonesia	Bali	Buyan	*Xiphophorus maculatus*	Dahruddin *et al* . 2016, Green et al. 1978
Indonesia	Bali	Buyan	*Xiphophorus hellerii*	Dahruddin *et al.* 2016
Indonesia	Bali	Danau Tamblir	*Anabas sp.*	Green *et al* . 1978
Indonesia	Bali	Tamblingan	*Barbodes microps*	Green *et al* . 1978
Indonesia	Bali	Tamblingan	*Channa striata*	Green *et al* . 1978
Indonesia	Bali	Tamblingan	*Clarias batrachus*	Green *et al* . 1978

Indonesia	Bali	Tamblingan	*Cyprinus carpio*	Green *et al* . 1978
Indonesia	Bali	Tamblingan	*Monopterus albus*	Green *et al* . 1978
Indonesia	Bali	Tamblingan	*Oreochromis mosambicus*	Green *et al* . 1978
Indonesia	Bali	Tamblingan	*Osteochilus vittatus*	Green *et al* . 1978
Indonesia	Bali	Tamblingan	*Poecilia reticulata*	Green *et al* . 1978
Indonesia	Bali	Tamblingan	*Rasbora sp.*	Green *et al* . 1978
Indonesia	Bali	Tamblingan	*Xiphophorus maculatus*	Green *et al* . 1978
Indonesia	Java	Rawa Pening	*Amphilophus citrinellus*	Dahruddin *et al.* 2016
Indonesia	Java	Rawa Pening	*Aplocheilus panchax*	Dahruddin *et al.* 2016
Indonesia	Java	Rawa Pening	*Barbonymus gonionotus*	Dahruddin *et al.* 2016
Indonesia	Java	Rawa Pening	*Channa striata*	Dahruddin *et al.* 2016
Indonesia	Java	Rawa Pening	*Clarias batrachus*	Dahruddin *et al.* 2016
Indonesia	Java	Rawa Pening	*Ctenopharyngodon idella*	Dahruddin *et al.* 2016
Indonesia	Java	Rawa Pening	*Cyprinus carpio*	Hutarabat, 1986
Indonesia	Java	Rawa Pening	*Dermogenys pusilla*	Dahruddin *et al.* 2016
Indonesia	Java	Rawa Pening	*Gobiopterus brachypterus*	Dahruddin *et al.* 2016
Indonesia	Java	Rawa Pening	*Monopterus albus*	Dahruddin *et al.* 2016
Indonesia	Java	Rawa Pening	*Notopterus notopterus*	Dahruddin *et al.* 2016
Indonesia	Java	Rawa Pening	*Oreochromis mossambicus*	Dahruddin *et al.* 2016
Indonesia	Java	Rawa Pening	*Osphronemus goramy*	Goeltenboth and Kristyanto 1994
Indonesia	Java	Rawa Pening	*Osteochilus vittatus*	Dahruddin et al. 2016
Indonesia	Java	Rawa Pening	*Oxyeleotris marmorata*	Dahruddin et al. 2016
Indonesia	Java	Rawa Pening	*Parachromis managuensis*	Dahruddin et al. 2016
Indonesia	Java	Rawa Pening	*Paraneetroplus fenestratus*	Dahruddin et al. 2017
Indonesia	Java	Rawa Pening	*Paraneetroplus maculicauda*	Dahruddin et al. 2018
Indonesia	Java	Rawa Pening	*Puntius brevis*	Dahruddin et al. 2018
Indonesia	Java	Rawa Pening	*Barbodes lateristriata*	Dahruddin et al. 2018
Indonesia	Java	Rawa Pening	*Trichopodus pectoralis*	Dahruddin et al. 2018
Indonesia	Java	Rawa Pening	*Trichopsis vittata*	Dahruddin et al. 2018
Indonesia	Kalimantan	Melintang	*Anabas testudineus*	Haryono, 2006
Indonesia	Kalimantan	Melintang	*Barbichthys laevis*	Haryono, 2006
Indonesia	Kalimantan	Melintang	*Barbonymus collingwoodii*	Haryono, 2006
Indonesia	Kalimantan	Melintang	*Hemibagrus nemurus*	Haryono, 2006
Indonesia	Kalimantan	Melintang	*Macrognathus aculeatus*	Haryono, 2006
Indonesia	Kalimantan	Melintang	*Osteochilus kappenii*	Haryono, 2006
Indonesia	Kalimantan	Melintang	*Pangasius sp.*	Haryono, 2006
Indonesia	Kalimantan	Melintang	*Parachela oxygastroides*	Haryono, 2006
Indonesia	Kalimantan	Melintang	*Pristolepis fasciata*	Haryono, 2006
Indonesia	Kalimantan	Melintang	*Thynnichthys vaillanti*	Haryono, 2006
Indonesia	Kalimantan	Melintang	*Trichopodus pectoralis*	Haryono, 2006
Indonesia	Kalimantan	Melintang	*Trichopodus trichopterus*	Haryono, 2006
Indonesia	Kalimantan	Semayang	*Anabas testudineus*	Haryono, 2006
Indonesia	Kalimantan	Semayang	*Barbichthys laevis*	Haryono, 2006; Kurniawan and Subehi, 2016
Indonesia	Kalimantan	Semayang	*Barbonymus collingwoodii*	Haryono, 2006

Indonesia	Kalimantan	Semayang	*Barbonymus schwanenfeldii*	Kurniawan and Subehi, 2016
Indonesia	Kalimantan	Semayang	*Helostoma temminckii*	Haryono, 2006
Indonesia	Kalimantan	Semayang	*Hemibagrus nemurus*	Haryono, 2006; Payuk et al. 2016
Indonesia	Kalimantan	Semayang	*Macrognathus siamensis*	Payuk et al. 2016
Indonesia	Kalimantan	Semayang	*Osteochilus kappenii*	Haryono, 2006
Indonesia	Kalimantan	Semayang	*Osteochilus kelabau*	Kurniawan and Subehi, 2016
Indonesia	Kalimantan	Semayang	*Osteochilus melanopleurus*	Payuk et al. 2016
Indonesia	Kalimantan	Semayang	*Osteochilus repang*	Payuk et al. 2016
Indonesia	Kalimantan	Semayang	*Osteochilus vittatus*	Kurniawan and Subehi, 2016
Indonesia	Kalimantan	Semayang	*Osteochilus waandersii*	Kurniawan and Subehi, 2016
Indonesia	Kalimantan	Semayang	*Oxyeleotris marmorata*	Haryono, 2006
Indonesia	Kalimantan	Semayang	*Oxygaster anomalura*	Payuk et al. 2016
Indonesia	Kalimantan	Semayang	*Pangasius sp.*	Haryono, 2006
Indonesia	Kalimantan	Semayang	*Parachela oxygastroides*	Kurniawan and Subehi, 2016
Indonesia	Kalimantan	Semayang	*Pristolepis fasciata*	Haryono, 2006
Indonesia	Kalimantan	Semayang	*Pseudomystus stenomus*	Payuk et al. 2016
Indonesia	Kalimantan	Semayang	*Rasbora sp.*	Haryono, 2006
Indonesia	Kalimantan	Semayang	*Striuntius lineatus*	Kurniawan and Subehi, 2016
Indonesia	Kalimantan	Semayang	*Striuntius lineatus*	Kurniawan and Subehi, 2016
Indonesia	Kalimantan	Semayang	*Thynichthys vaillanti*	Payuk et al. 2016; Haryono, 2006
Indonesia	Kalimantan	Semayang	*Trichopodus trichopterus*	Haryono, 2006
Indonesia	Kalimantan	Riam Kanan	*Cyprinus carpio*	Hardjamulia and Suwignyo, 1988
Indonesia	Kalimantan	Riam Kanan	*Colossoma macropomum*	Rahman et al. 2017
Indonesia	Kalimantan	Riam Kanan	*Hampala macrolepidota*	Hardjamulia, A. and Suwignyo, P., 1988
Indonesia	Kalimantan	Riam Kanan	*Hemibagrus nemurus*	Hardjamulia, A. and Suwignyo, P., 1988
Indonesia	Kalimantan	Riam Kanan	*Ophicephalus seiatus*	Hardjamulia, A. and Suwignyo, P., 1988
Indonesia	Kalimantan	Riam Kanan	*Puntius gonionatus*	Hardjamulia, A. and Suwignyo, P., 1988
Indonesia	Kalimantan	Riam Kanan	*Oreochromis niloticus*	Hardjamulia, A. and Suwignyo, P., 1988
Indonesia	Kalimantan	Riam Kanan	*Osphronemus goramy*	Tanjung et al. 2013
Indonesia	Sulawesi	Matano	*Ophisternon bengalense*	Herder et al. 2012
Indonesia	Sulawesi	Matano	*Anabas testudineus*	Versteegh, D. 2010
Indonesia	Sulawesi	Matano	*Anguilla marmorata*	Versteegh, D. 2010
Indonesia	Sulawesi	Matano	*Anguilla nebulosa*	Versteegh, D. 2010
Indonesia	Sulawesi	Matano	*Aplocheilus panchax*	Versteegh, D. 2010

Indonesia	Sulawesi	Matano	*Channa striata*	Versteegh, D. 2010
Indonesia	Sulawesi	Matano	*Colossoma macropomum*	Herder et al. 2012
Indonesia	Sulawesi	Matano	*Clarias batrachus*	Versteegh, D. 2010
Indonesia	Sulawesi	Matano	*Cyprinus carpio*	Versteegh, D. 2011
Indonesia	Sulawesi	Matano	*Nomorhamphus megarrhamphus*	Versteegh, D. 2012
Indonesia	Sulawesi	Matano	*Nomorhamphus weberi*	Versteegh, D. 2013
Indonesia	Sulawesi	Matano	*Glossogobius matanensis*	Nasution, 2016
Indonesia	Sulawesi	Matano	*Monopterus albus*	Versteegh, D. 2010
Indonesia	Sulawesi	Matano	*Mugilogobius adeiae*	Nasution, 2016
Indonesia	Sulawesi	Matano	*Mugilogobius latifrons*	Nasution, 2016
Indonesia	Sulawesi	Matano	*Nomorhamphus brembachi*	Nasution, 2016
Indonesia	Sulawesi	Matano	*Oreochromis mosambicus*	Herder et al. 2012
Indonesia	Sulawesi	Matano	*Oryzias mamoratus*	Versteegh, 2010
Indonesia	Sulawesi	Matano	*Oryzias matanensis*	Nasution, 2016
Indonesia	Sulawesi	Matano	*Paratherina wolterecki*	Versteegh, 2010
Indonesia	Sulawesi	Matano	*Poecilia reticulata*	Herder et al. 2012
Indonesia	Sulawesi	Matano	*Pseudotropheus cyaneorhabdos*	Herder et al. 2012
Indonesia	Sulawesi	Matano	*Pterygoplichthys pardalis*	Herder et al. 2012
Indonesia	Sulawesi	Matano	*Mugilogobius sarasinorum*	Versteegh, 2010
Indonesia	Sulawesi	Matano	*Telmatherina abendanoni*	Nasution, 2016; Kurniawan and Subehi, 2016
Indonesia	Sulawesi	Matano	*Telmatherina antoniae*	Nasution, 2016; Kurniawan and Subehi, 2016
Indonesia	Sulawesi	Matano	*Telmatherina bonti*	Nasution, 2016; Kurniawan and Subehi, 2016
Indonesia	Sulawesi	Matano	*Telmatherina celebensis*	Versteegh, 2010
Indonesia	Sulawesi	Matano	*Telmatherina obscura*	Nasution, 2016; Kurniawan and Subehi, 2016
Indonesia	Sulawesi	Matano	*Telmatherina opudi*	Nasution, 2016; Kurniawan and Subehi, 2016
Indonesia	Sulawesi	Matano	*Telmatherina prognatha*	Kurniawan and Subehi, 2016
Indonesia	Sulawesi	Matano	*Telmatherina sarasinorum*	Nilawati et al. 2010; Nasution, 2016; Kurniawan and Subehi, 2016
Indonesia	Sulawesi	Matano	*Telmatherina wahjui*	Nasution, 2016; Kurniawan and Subehi, 2016
Indonesia	Sulawesi	Matano	*Trichopodus pectoralis*	Versteegh, 2010
Indonesia	Sumatra	Laut Tawar	*Anguilla marmorata*	Muchlisin *et al* . 2010
Indonesia	Sumatra	Laut Tawar	*Channa gachua*	Muchlisin *et al* . 2010
Indonesia	Sumatra	Laut Tawar	*Channa striata*	2009; Muchlisin *et al* .

Indonesia	Sumatra	Laut Tawar	*Clarias gariepinus*	Muchlisin *et al.* 2010
Indonesia	Sumatra	Laut Tawar	*Clarias sp.*	Muchlisin *et al* . 2010
Indonesia	Sumatra	Laut Tawar	*Ctenopharyngodon idella*	Muchlisin et al. 2009; Muchlisin and Azizah, 2009; Muchlisin, 2012
Indonesia	Sumatra	Laut Tawar	*Cyprinus carpio*	Muchlisin et al. 2009; Muchlisin and Azizah, 2009; Muchlisin et al. 2010; Muchlisin, 2012
Indonesia	Sumatra	Laut Tawar	*Pterygoplichthys pardalis*	Muchlisin *et al* . 2009
Indonesia	Sumatra	Laut Tawar	*Homaloptera sp.*	Muchlisin *et al* . 2009
Indonesia	Sumatra	Laut Tawar	*Oreochromis mossambicus*	Muchlisin *et al.* 2009; Muchlisin and Azizah, 2009; Muchlisin, 2012
Indonesia	Sumatra	Laut Tawar	*Oreochromis niloticus*	Muchlisin *et al.* 2009; Muchlisin and Azizah, 2009; Muchlisin *et al.* 2010; Muchlisin, 2012
Indonesia	Sumatra	Laut Tawar	*Osteochilus kahajanensis*	Muchlisin *et al* . 2009
Indonesia	Sumatra	Laut Tawar	*Poropuntius tawarensis*	2009; Muchlisin *et al.*
Indonesia	Sumatra	Laut Tawar	*Puntius brevis*	2009; Muchlisin *et al* .
Indonesia	Sumatra	Laut Tawar	*Rasbora meinkeni*	Lumbantobing, 2010
Indonesia	Sumatra	Laut Tawar	*Rasbora sp.*	Muchlisin and Azizah,
Indonesia	Sumatra	Laut Tawar	*Rasbora tawarensis*	Muchlisin and Azizah,
Indonesia	Sumatra	Laut Tawar	*Xiphophorus hellerii*	Muchlisin *et al* . 2009; Muchlisin and Azizah, 2009; Muchlisin *et al* . 2010; Muchlisin, 2012
Indonesia	Sumatra	Laut Tawar	*Xiphophorus maculatus*	2009; Muchlisin, 2012.
Indonesia	Sumatra	Laut Tawar	*Trichopodus trichopterus*	Muchlisin *et al* . 2009
Indonesia	Sumatra	Singkarak	*Anabas testudeneus*	Oktavia and Faoziyah, 2016; Mardiah *et al.* 2016
Indonesia	Sumatra	Singkarak	*Cyclocheilichthys armatus*	Oktavia and Faoziyah, 2016
Indonesia	Sumatra	Singkarak	*Barbodes belinka*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Barbonymus schwanenfeldii*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Channa lucius*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Channa striata*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Clarias batrachus*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Cyclocheilichthys apogon*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Cyclocheilichthys armatus*	Wetlands International, Indonesia, Danau Singkarak; Mardiah et al. 2016

Indonesia	Sumatra	Singkarak	*Cyprinus carpio*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Glyptothorax platypogonoides*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Gobiopterus brachypterus*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Hampala macrolepidota*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Hampala bimaculata*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Hemibagrus nemurus*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Homaloptera gymnogaster*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Mastacembelus erythrotaenia*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Mastacembelus unicolor*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Mystacoleucus padangensis*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Hemibagrus planiceps*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Oreochromis mossambicus*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Oreochromis niloticus*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Osphronemus goramy*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Osteochilus kappenii*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Osteochilis vittatus*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Osteochilus waandersii*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Psilotris sp*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Rasbora argyrotaenia*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Rasbora jacobsoni*	Wetlands International, Indonesia, Danau Singkarak

Indonesia	Sumatra	Singkarak	*Rasbora spilotaenia*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Arothron mappa*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Pao palembangensis*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Tor douronensis*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Singkarak	*Tor tambroides*	Wetlands International, Indonesia, Danau Singkarak
Indonesia	Sumatra	Singkarak	*Trichopodus trichopterus*	Oktavia and Faoziyah, 2016; Mardiah et al. 2016
Indonesia	Sumatra	Toba	*Rasbora tobana*	Fishbase, Rasbora tobana
Indonesia	Sumatra	Toba	*Anabas testudineus*	Fishbase Danau Toba, Wijopriono et al. 2010
Indonesia	Sumatra	Toba	*Aplocheilus panchax*	Fishbase Danau Toba
Indonesia	Sumatra	Toba	*Barbodes binotatus*	Fishbase Danau Toba
Indonesia	Sumatra	Toba	*Barbonymus gonionotus*	Fishbase Danau Toba; Wijopriono et al. 2010
Indonesia	Sumatra	Toba	*Barbonymus schwanenfeldii*	Fishbase Danau Toba
Indonesia	Sumatra	Toba	*Betta imbellis*	Fishbase Danau Toba
Indonesia	Sumatra	Toba	*Betta taeniata*	Fishbase Danau Toba
Indonesia	Sumatra	Toba	*Channa gachua*	Fishbase Danau Toba
Indonesia	Sumatra	Toba	*Channa striata*	Fishbase Danau Toba
Indonesia	Sumatra	Toba	*Clarias batrachus*	Fishbase, 2017; Wijopriono et al. 2010
Indonesia	Sumatra	Toba	*Clarias nieuhofii*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Ctenopharyngodon idella*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Cyprinus carpio*	Fishbase, 2017; Wijopriono et al. 2010
Indonesia	Sumatra	Toba	*Danio albolineatus*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Hampala macrolepidota*	Wijopriono et al. 2010
Indonesia	Sumatra	Toba	*Homalopterula gymnogaster*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Poecilia reticulata*	Fishbase, 2017; Wijopriono et al. 2010
Indonesia	Sumatra	Toba	*Monopterus albus*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Mystacoleucus padangensis*	Panjaitan, 2010; Wijopriono et al. 2010
Indonesia	Sumatra	Toba	*Nemacheilus pfeifferae*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Nemacheilus fasciatus*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Neolissochilus thienemanni*	Fishbase, 2017; Saragih and Sunito, 2001.
Indonesia	Sumatra	Toba	*Oreochromis mossambicus*	Fishbase, 2017; Wijopriono et al. 2010
Indonesia	Sumatra	Toba	*Oreochromis niloticus*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Osphronemus goramy*	Fishbase, 2017

Indonesia	Sumatra	Toba	*Osteochilus vittatus*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Oxyeleotris marmorata*	Wijopriono et al, 2010
Indonesia	Sumatra	Toba	*Barbodes binotatus*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Rasbora jacobsoni*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Tor tambra*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Tor duoronensis*	Wijopriono et al. 2010
Indonesia	Sumatra	Toba	*Trichopodus pectoralis*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Trichopodus trichopterus*	Fishbase, 2017
Indonesia	Sumatra	Toba	*Xiphophorus hellerii*	Fishbase, 2017
Malaysia	Perak	Chenderoh	*Barbonymus gonionotus*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Barbonymus schwanenfeldii*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Channa micropeltes*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Channa striata*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Chitala chitala*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Chitala lopis*	Kah-Wai and Ali, 2000
Malaysia	Perak	Chenderoh	*Cichla ocellaris*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Ctenopharyngodon idella*	Kah-Wai and Ali, 2000
Malaysia	Perak	Chenderoh	*Cyclocheilichthys apogon*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Cyclocheilichthys armatus*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Cyclocheilichthys heteronema*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Epalzeorhynchos spp*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Hampala macrolepidota*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Hemibagrus nemurus*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Hypophthalmichthys molitrix*	Kah-Wai and Ali, 2000
Malaysia	Perak	Chenderoh	*Hypophthalmichthys nobilis*	Kah-Wai and Ali, 2000
Malaysia	Perak	Chenderoh	*Hypsibarbus wetmorei*	Kah-Wai and Ali, 2000
Malaysia	Perak	Chenderoh	*Labiobarbus fasciatus*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Labiobarbus leptocheilus*	Kah-Wai and Ali, 2000
Malaysia	Perak	Chenderoh	*Labiobarbus lineatus*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Leptobarbus hoevenii*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Mastacembelus erythrotaenia*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Mastacembelus favus*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Mystacoleucus marginatus*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Mystus castaneus*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Notopterus notopterus*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Oreochromis sp.*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Osphronemus goramy*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Osteochilus melanopleurus*	Kah-Wai and Ali, 2000
Malaysia	Perak	Chenderoh	*Osteochilus microcephalus*	Hashim *et al* . 2012

Malaysia	Perak	Chenderoh	*Osteochilus vittatus*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Oxyeleotris marmorata*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Oxygaster anomalura*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Pao leiurus*	Kah-Wai and Ali, 2000
Malaysia	Perak	Chenderoh	*Poropuntius deauratus*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Pristolepis fasciata*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Pristolepis grootii*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Pseudolais micronemus*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Puntigrus partipentazona*	Kah-Wai and Ali, 2000
Malaysia	Perak	Chenderoh	*Puntioplites bulu*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Rasbora sumatrana*	Kah-Wai and Ali, 2000
Malaysia	Perak	Chenderoh	*Rasbora tornieri*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Thynnichthys thynnoides*	Kah-Wai and Ali, 2000; Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Trichopodus trichopterus*	Hashim *et al* . 2012
Malaysia	Perak	Chenderoh	*Xenentodon canciloides*	Hashim *et al* . 2012

Appendix 4: References

Budiasa, I. W., Santosa, I. G. N., Ambarawati, I. G. A. A., Suada, I. K., Sunarta, I. N., Shchegolkova, N. 2018. Feasibility study and carrying capacity of Lake Batur ecosystem to preserve tilapia fish farming in Bali, Indonesia. BIODIVERSITAS. ISSN: 1412-033X Volume 19, Number 2, March 2018 E-ISSN: 2085-4722 Pages: 613-620

Dahruddin, H., Hutama, A., Busson, F., Sauri, S., Hanner, R., Keith, P., Hadiaty, R. and Hubert, N., 2016. Revisiting the ichthyodiversity of Java and Bali through DNA barcodes: taxonomic coverage, identification accuracy, cryptic diversity and identification of exotic species. Molecular Ecology Resources, 17(2), pp.288-299.

Fishbase. 2017. Species in Toba. Accessed at: http://fishbase.org/trophiceco/FishEcoList.php?ve_code=547

Green, J., Corbet, S.A., Watts, E. and Lan, O.B., 1978. Ecological studies on Indonesian lakes. The montane lakes of Bali. Journal of Zoology, 186(1), pp.15-38.

Hashim, Z.H., Zainuddin, R.Y., Shah, A.S.R.M., Sah, S.A.M., Mohammad, M.S. and Mansor, M., 2012. Fish checklist of Perak River, Malaysia. Check List, 8(3), pp.408-413.

Hutarabat, J., Syarani, L. and Smith, M.A.K., 1986. Use of freshwater hyacinth Eichhornia crassipes in cage culture in Lake Rawa Pening, Central Java. In 1. Asian Fisheries Forum, Manila (Philippines), 26-31 May 1986.

Kah-Wai, K. and Ali, A.B., 2000, February. Chenderoh Reservoir, Malaysia: Fish community and artisanal fishery of a small mesotrophic tropical reservoir. In ACIAR PROCEEDINGS (pp. 167-

Kartamihardja, E.S., 2012. STOCK ENHANCEMENT IN INDONESIAN LAKE AND RESERVOIRS FISHERIES. Indonesian Fisheries Research Journal, 18(2), pp.91-100.

Mardiah, A., Azrita, and Syandri, H. FISH DIVERSITY OF THE SINGKARAK LAKE, INDONESIA: PRESENT STATUS AND CONSERVATION NEEDS. Proceedings of the 16th World Lake Conference.

Rahman A, Sentosa A A, Wijaya D. 2012. Sebaran ukuran dan kondisi ikan zebra Amatitlania nigrofascia (Günther, 1867) di Danau Beratan, Bali. Jurnal Iktiologi Indonesia.12 (2):135-145

Restu, I. W., Kartika, G. R. A., Pratiwi, M. A. 2016. POTENTIAL IDENTIFICATION OF FLORA AND FAUNA LAKE BUYAN AS BASIS FOR TOURISM DEVELOPMENT STRATEGY BASED ON AQUATIC ECOSYSTEMS. Proceedings of the 16th World Lake Conference.

Sentosa, A.A. and Wijaya, D. 2012. PEMANFAATAN MAKANAN ALAMI OLEH IKAN-IKAN DOMINAN DI DANAU BATUR, PROVINSI BALI. Seminar Nasional Tahunan IX Hasil Penelitian Perikanan dan Kelautan, 14 Juli 2012. UTILIZATION OF NATURAL FOOD BY DOMINANT FISH IN BATUR LAKE, BALI PROVINCE. Annual National Seminar IX Fisheries and Marine Research Results, July 14, 2012

Suryaningtyas, E. U., Ulinuha, D. EFFECT OF SEASONAL CHANGES ON SPATIAL DISTRIBUTION OF BACTERIAL PATHOGENS IN TILAPIA (Oreochromis niloticus) IN LAKE BATUR. Proceedings of the 16th World Lake Conference.

Whitten, T., Soeriaatmadja, R.E. and Afiff, S.A., 1996. Ecology of Java & Bali (Vol. 2). Oxford University Press.

Appendix 5

Primer and primer-index sequences

Primer information

Primer name	Gene	Animal Group	F Primer Sequence	F [bp]	F and R primer index sequence	Index [bp]	R Primer sequence	R [bp]	Index [bp]	Target size	Insert total size	Indexed total size
Leray 1	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NAACAAC	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 2	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NAACCGA	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 3	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NGCTTAA	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 4	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NGTGTAT	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 5	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NAACGCT	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 6	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NCTAAGC	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 7	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NGTTACA	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 8	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NAAGACA	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 9	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NACGTGA	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 10	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NTCTGCA	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 11	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NAAGCAT	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 12	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NCCATTC	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 13	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NAGACTC	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 14	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NATTATC	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 15	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NTGTGAC	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 16	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NAAGGTC	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 17	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NACTCCT	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 18	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NGTGGTA	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 19	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NTATTAT	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 20	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NCTCCAT	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 21	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NGGTCTA	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 22	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NAATAGT	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 23	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NCCGAAT	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 24	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NCAACAC	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 25	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NTTGTCC	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 26	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NTAAGGC	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 27	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NCCTAGA	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 28	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NAATGAA	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 29	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NTGAGTA	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504

Leray 30	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NATAGAC	7	TAXACYTCXGGRTGXCCRAARAAYCA	26	7	313	379	504
Leray 31	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNAGAAGA	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 32	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNTCTTGC	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 33	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNTTCAGA	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 34	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNGTACGA	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 35	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNAATTCC	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 36	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNTGCAAT	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 37	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNCAATGT	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 38	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNACAACC	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 39	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNATATTA	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 40	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNTACCTC	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 41	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNCGAGAT	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 42	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNTATATA	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 43	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNTGCTCA	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 44	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNCACTAA	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 45	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNAGATCT	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 46	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNTGTCGT	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 47	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNTAACCT	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 48	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNACAGGT	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 49	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNTGGATC	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 50	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNTGGCAA	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 51	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNCAAGCA	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 52	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNATCTGC	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 53	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNCGTACT	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 54	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNACACAA	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 55	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNACCATA	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 56	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNGTTGGT	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 57	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNCAGCTA	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 58	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNACCTAT	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 59	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNAGGTAC	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Leray 60	COI	Metazoa	GGWACWGGWTGAACWGTWTAYCCYCC	26	NNGTTCAC	8	TAXACYTCXGGRTGXCCRAARAAYCA	26	8	313	381	506
Valentini 1	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNAACAAC	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 2	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNAACCGA	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 3	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNCCGGAA	9	CTTCCGGTACACTTACCATG	20	9	63	118	243

Valentini 4	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNAGTGTT	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 5	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNCCGCTG	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 6	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNAACGCG	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 7	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNGGCTAC	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 8	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNTTCTCG	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 9	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNTCACTC	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 10	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNGAACTA	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 11	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNCCGTCC	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 12	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNAAGACA	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 13	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNCGTGCG	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 14	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNGGTAAG	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 15	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNATAATT	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 16	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNCGTCAC	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 17	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNTTGAGT	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 18	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNAAGCAG	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 19	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNTTGCAA	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 20	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNCACGTA	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 21	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNTAACAT	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 22	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNTGCGTG	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 23	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNGGTCGA	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 24	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNCACTCT	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 25	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNCTTGGT	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 26	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNTCCAGC	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 27	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNACTTCA	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 28	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNGCGAGA	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 29	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNTGGAAC	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 30	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNGTACAC	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 31	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNNAAGTGT	9	CTTCCGGTACACTTACCATG	20	9	63	118	243
Valentini 32	12S	Teleosti	ACACCGCCCGTCACTCT	17	NNTCTTGG	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 68	12S	Teleosti	ACACCGCCCGTCACTCT	17	ATCGCAGC	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 69	12S	Teleosti	ACACCGCCCGTCACTCT	17	TGAGCAGC	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Valentini 70	12S	Teleosti	ACACCGCCCGTCACTCT	17	ACGACAGC	8	CTTCCGGTACACTTACCATG	20	8	63	116	241
Taylor 1	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TCTGCGAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 2	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ATCAGCAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276

Taylor 3	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ATACAGTC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 4	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ATCATATC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 5	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TGCGATGC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 6	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ATATACGC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 7	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ATCGCAGC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 8	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TATACTAC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 9	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ACTACGAC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 10	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	AGCATCAC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 11	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ATAGAGAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 12	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TATCAGAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 13	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ACGCAGAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 14	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ACAGTCAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 15	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TCTATCAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 16	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TAGTGCAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 17	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TGCTACAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 18	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	AGTGACAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 19	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ACTGTGTC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 20	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TACATGTC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 21	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TCAGTGCG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 22	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	GTAGCAGA	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 23	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ATTCACAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 24	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ATTCCATA	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 25	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TGGCCGAT	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 26	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ATGCATAC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 27	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ATGCCGCA	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 28	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TAACTACT	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 29	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ACACTACG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 30	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	AGACCATC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 31	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	GCCGAGAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 32	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TAAGTCAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 33	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ACAGGCAG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 34	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ACAGAGTC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 35	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TCAGTATC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 36	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TAAGGTGC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276

Taylor 37	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TGAGCTAC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 38	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	AGAGTGAC	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 39	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ACTCTGTG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 40	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TATCCATG	8	GCTGTTATCCCTAGGGTAACT	21	8	95	151	276
Taylor 41	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	GGCTCAT	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 42	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	CATGCTC	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 43	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TCATCGG	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 44	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	CATCTAT	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 45	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	GTCACAG	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 46	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TATGCAT	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 47	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	GCGAGAC	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 48	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	GCATCAC	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 49	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	AGTGTCC	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 50	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ATGCGTC	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 51	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	CCGGTCC	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 52	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TATCTCC	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 53	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	TGTCAGT	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 54	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	CCTGCAG	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 55	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	GGCAGTG	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 56	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	CGTTGCC	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 57	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	AGGTCGT	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 58	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	ACGTCAG	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 59	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	CAGACAC	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274
Taylor 59	16S	Mammalia	CGGTTGGGGTGACCTCGGA	19	GCACGTG	7	GCTGTTATCCCTAGGGTAACT	21	7	95	149	274

All removed reads

OTU	Taxonomy	Query	Identity	Marker	Reason
100	No hits	91	80	12S	Chimera
101	No hits	57	84	12S	Chimera
110	No hits	24	93	12S	Query < 55
122	No hits	100	79	12S	Bacteria
125	No hits	76	88	12S	Bacteria
128	No hits	93	78	12S	Chimera
21	No hits	29	94	12S	Query < 55
40	No hits	41	93	12S	Query < 55
42	No hits	100	87	12S	Bacteria
95	No hits	57	89	12S	Chimera
97	No hits	31	91	12S	Query < 55
127	Homo sapiens chromosome 7	100	100	12S	Contaminant
1	Homo sapiens haplogroup	100	100	12S	Contaminant
132	Homo sapiens haplogroup	57	100	12S	Contaminant
64	Homo sapiens chromosome 8	100	100	12S	Contaminant
38	Gallus	100	100	12S	Contaminant
111	Dicentrarchus labrax	100	100	12S	+ve control
118	Dicentrarchus labrax	100	100	12S	+ve control
13	Dicentrarchus labrax	100	100	12S	+ve control
17	Dicentrarchus labrax	100	100	12S	+ve control
99	Ctenolabrus rupestris	100	100	12S	+ve control
78	Labrus merula	100	100	12S	+ve control
48	Sparus aurata	100	100	12S	+ve control
66	Mustelus manazo	100	98	12S	+ve control
49	Raja clavata	100	98	12S	+ve control
80	Raja clavata	100	98	12S	+ve control
105	Psetta maxima	100	100	12S	+ve control
96	Scyliorhinus canicula	100	95	12S	+ve control
86	Micromesistius poutassou	100	100	12S	+ve control
36	Labrus mixtus	85	100	12S	+ve control
43	Anarhichas lupus	100	100	12S	+ve control
52	Scomber scombrus	100	100	12S	+ve control
61	Anguilla anguilla	100	100	12S	+ve control
35	Trichogaster microlepis	100	85	12S	+ve control
26	Pethia cumingii	100	100	12S	+ve control
27	Puntius titteya	100	100	12S	+ve control
29	Microctenopoma ansorgii	100	100	12S	+ve control
4	Channa striata	93	100	12S	+ve control
7	Balantiocheilos melanopterus	100	100	12S	+ve control
123	Barbodes lateristriga	100	95	12S	+ve control
3	Barbodes lateristriga	100	100	12S	+ve control
73	Barbodes lateristriga	96	100	12S	+ve control
6	Rasbora borapetensis	100	92	12S	+ve control

71	Probarbus jullieni	100	100	12S	+ve control
8	Anabas testudineus	100	100	12S	+ve control
5	No hits			12S	+ve control
9	Eutaeniichthys gilli	100	84	12S	+ve control
11	No hits	90	95	12S	+ve control
117	No hits	71	93	12S	+ve control
133	No hits	100	92	12S	+ve control
134	No hits			12S	+ve control
135	No hits			12S	+ve control
136	No hits			12S	+ve control
137	No hits			12S	+ve control
138	No hits			12S	+ve control
139	No hits			12S	+ve control
140	No hits			12S	+ve control
141	No hits			12S	+ve control
142	No hits			12S	+ve control
143	No hits			12S	+ve control
145	No hits			12S	+ve control
146	No hits			12S	+ve control
149	No hits			12S	+ve control
150	No hits			12S	+ve control
55	No hits	100	91	16S	+ve control
81	No hits			16S	+ve control
75	No hits			16S	+ve control
74	No hits			16S	+ve control
71	No hits			16S	+ve control
60	No hits			16S	+ve control
57	No hits			16S	+ve control
51	Not assigned			16S	+ve control
80	Cetacea	100	91	16S	+ve control
73	Cetacea	97	93	16S	+ve control
70	Cetacea	78	92	16S	+ve control
56	Cetacea	95	94	16S	+ve control
69	Odontoceti	97	97	16S	+ve control
68	Odontoceti	100	99	16S	+ve control
2	Phocoena phocoena	100	100	16S	+ve control
52	Hipposideros ridleyi	88	88	16S	+ve control
54	Laurasiatheria	100	87	16S	+ve control
61	Rhinopoma	95	90	16S	+ve control
31	Bufo bufo	100	100	16S	+ve control
35	Elephas maximas	100	100	16S	+ve control
5	Microtus sp.	100	95	16S	+ve control
27	Myodes glareolus	100	100	16S	+ve control
9	Oryctolagus cuniculus	100	100	16S	+ve control
24	Aonyx cinerea	100	100	16S	+ve control

64	Giraffa camelopardalis	100	100	16S	+ve control
15	Rutilus rutilus	100	100	16S	+ve control
47	Catarrhini	100	99	16S	Contaminant
49	Homininae	100	99	16S	Contaminant
34	Homininae	100	100	16S	Contaminant
28	Homininae	100	100	16S	Contaminant
21	Homo sapiens	100	100	16S	Contaminant
88	Homo sapiens	100	100	16S	Contaminant
78	Homo sapiens	100	100	16S	Contaminant
59	Homo sapiens	100	100	16S	Contaminant
32	Homo sapiens	100	100	16S	Contaminant
1	Homo sapiens	100	100	16S	Contaminant
44	Pseudoryx nghetinhensis	100	96	16S	Contaminant
3	Pseudoryx nghetinhensis	100	100	16S	Contaminant
9	Acaudina molpadioides	50	76	68516	Query < 55
11	Bilateria	47	79	53052	Query < 55
15	Troglodytes aedon	51	78	32356	Query < 55
28	Gelidium omanense	15	88	18307	Query < 55
38	Microplitis	50	82	14723	Query < 55
41	Rhopaea magnicornis	12	95	14532	Query < 55
58	Eumetazoa	52	86	8837	Query < 55
63	Actitis macularia	30	84	8067	Query < 55
64	Melosira ambigua2	100	98	25577	Present in -ve
77	Tricholoma matsutake	35	83	6315	Query < 55
78	Pseudopediastrum boryanum	50	83	10468	Query < 55
80	Bilateria2	32	81	6414	Query < 55
83	Dorvilleidae sp.	8	100	10804	Query < 55
98	Neoptera	50	83	4728	Query < 55
100	Pyramimonas parkeae	43	85	4665	Query < 55
114	Protostomia4	5	85	6405	Query < 55
141	Homo sapiens	100	100	5226	Contaminant
142	Brachionus dimidiatus	22	89	2872	Query < 55
143	Corallina	16	89	5612	Query < 55
164	Bilateria3	52	82	2298	Query < 55
176	Pedinomonas minor	89	88	2063	Present in -ve
183	Vireo olivaceus	53	81	1867	Query < 55
203	Protostomia8	44	80	1527	Query < 55
206	Eukaryota34	52	80	1490	Query < 55
214	Microbacterium foliorum	18	83	1429	Query < 55
222	Bilateria4	51	80	1366	Query < 55
223	Diptera	100	92	1702	Present in -ve
243	Opisthokonta6	23	88	1171	Query < 55
248	Dermogenys pusilla	23	84	1131	Query < 55
259	Pseudoryx nghetinhensis	100	100	1064	Contaminant
261	Sigmurethra	22	83	1041	Query < 55

266	Eukaryota42	51	78	999	Query < 55
267	Bilateria5	49	78	999	Query < 55
274	Bilateria6	27	82	975	Query < 55
281	Eukaryota43	32	83	917	Query < 55
290	Diptera2	30	84	868	Query < 55
307	Eukaryota47	53	78	772	Query < 55
330	Protostomia10	35	86	630	Query < 55
344	Protostomia11	35	86	570	Query < 55
348	Arthropoda	28	89	563	Query < 55
358	Eukaryota56	53	78	522	Query < 55
375	Holometabola3	50	82	459	Query < 55
376	Eukaryota2	19	90	757	Query < 55
400	Eumetazoa5	22	89	414	Query < 55
401	Eumetazoa6	52	72	580	Query < 55
407	Coleoptera	30	83	397	Query < 55
414	Elaenia flavogaster	48	82	378	Query < 55
425	Bilateria9	51	79	361	Query < 55
428	Eukaryota70	53	82	357	Query < 55
443	Gammarus balcanicus	27	84	335	Query < 55
450	Neogovea sp.	12	95	324	Query < 55
451	Bilateria10	37	80	322	Query < 55
459	Noctuoidea	51	80	308	Query < 55
465	Hypoaspis sp.	25	88	300	Query < 55
467	Eukaryota	51	80	298	Query < 55
475	Microhedyle glandulifera	50	81	282	Query < 55
497	Bilateria11	27	89	243	Query < 55
515	Pagurixus nomurai	47	79	215	Query < 55
517	Eukaryota90	50	76	213	Query < 55
544	Psoroptidia	98	86	191	Present in -ve
564	Bayerotrochus	50	83	175	Query < 55
577	Neoptera5	27	84	166	Query < 55
583	Siphonaria	48	82	162	Query < 55
589	Pheidole	38	83	152	Query < 55
590	Neoptera6	47	79	151	Query < 55
591	Chrysaora chinensis	53	79	147	Query < 55

Appendix 7

Full bioinformatic pipeline

Commands are given in italics, and commands in bold require specific information not given in this text dependent upon unique file names / numbers etc.

1. Create the primer information file using three separate files created in a text editor. File1 contains the primer name, file2 contains the forward primer sequence, and file3 contains the reverse primer sequence. Merge these files together to create one new file using the paste command.

```
paste file1.txt file2.txt file3.txt > Primer_Info_Filename.txt
```

2. Create the tag information file using two separate files created in a text editor. File1 contains the tag sequence, file2 contains the tag number (in the format Tag1, Tag2, Tag3… etc)

```
paste file1.txt file2.txt > Tag_Info_Filename.txt
```

3. Create the PSInfo file using four separate files created in a text editor. File 1 contains the Sample name, file2 contains the forward tag number, file3 contains the reverse tag number, and file 4 contains the pool in which the sample was placed. Then add the word 'Tag' in front of the Tag numbers to make files accessible for DAMe. * check why the file prep works this way, why do we need a separate primer and tag file if PS Info files have the Tag information already in them? What does PS Info stand for?

```
Paste file1.txt file2.txt file3.txt > PS_Info_Filename.txt
awk '{print $1"\tTag"$2"\tTag"$3"\t"$4}' PS_Info_Filename.txt
```

Accessing and viewing raw sequencing files

4. Download sequences from link provided by Copenhagen Sequencing centre

```
wget –r data link
```

5. Load the programme FastQC and look at the help file

```
module load fastqc/v0.11.5

fastqc –h
```

6. Run fastqc on file of interest. This creates a range of new fastq files.

```
fastqc file
```

7. The file: fastq.gz must be unzipped using the function: gunzip.

```
gunzip file
```

8. Use FileZilla application to download files from the server to the computer hard drive, and open fastqc.html files to view graphs and data summaries.

Adapter removal and paired-end read merging

9. Load the programme AdapterRemoval and look at the help file.

```
module load AdapterRemoval

AdapterRemoval –h
```

10. Make new directories for each sequencing pool to store the new files which will be created.

```
P=3
for i in `seq 1 $P`
do
mkdir pool${i}
done
```

11. Run the AdapterRemoval function on the raw sequencing files containing one fastqc file with the forward read (R1), and one fastqc file with the reverse read (R2). This removes

the adapters. Add variables to filter for minimum length (minlength) (shift) (basename) (trimns) (trimqualities) (qualitybase) (minquality) (minalignmentlength) (collapse)

```
AdapterRemoval --file1 filename.fastq --file2 filename.fastq --minlength 50 --shift 5 --basename pool1_merged --trimns --trimqualities --qualitybase 33 --minquality 28 --minalignmentlength 20 –collapse
```

12. Create a merged fastq file either one at a time:

```
cat poolnumber_merged.collapsed poolnumber_merged.collapsed.truncated > Poolnumber_merged.fastq
```

Or many at a time:

```
P = 3
for i in `seq 1 $P`
do
cd /directory path/primer file/pool${i}
cat pool${i}_merged.collapsed pool${i}_merged.collapsed.truncated > Pool${i}_merged.fastq
cd ../
done
```

13. View FastQC information on new merged files one at a time:

```
fastqc file_merged.fastq
```

Or many at a time:

```
P=3
for i in `seq 1 $P`
do
cd /directory path/primer file/pool${i}
fastqc Pool${i}_merged.fastq
```

```
cd ../
done
```

Sorting sequence information by sample information

14. Load the necessary programmes and functions and view their help files:

```
module load python/v2.7.12
module load DAMe/v0.9
DAMEe -h
module load sort.py
sort.py -h
```

15. Sort the merged fasta files according to the primers and tags, and view the number of erroneous sequences that occur which have an error in the primer, tag or no barcode amplification. This command also makes various files explain here

```
P=3
for i in `seq 1 $P`
do
cd /directory path/pool${i}
sort.py -fq Pool${i}_merged.fastq -p /DirectoryPath/PrimerFile.txt -t
/DirectoryPath/TagsFile.txt
done
```

Example of the output:

```
Number of erroneous sequences in file Pool1_merged.fastq (with errors in the sequence of
primer or tags, or no barcode amplified): 48577
   No sequence between primers        : 5
   Tags pair not found                : 23674
   F primer found, R' primer not found : 6127
   R primer found, F' primer not found : 5132
```

Neither F nor R primer found : 13639

Number of valid tag pairs found : 258816

F-R' barcodes found : 134588

R-F' barcodes found : 124228

Tags are not all the same length.

Among the tags with no mismatches, the longest one will be retained.

16. Make a sorted summary counts file from the summary counts file.

```
P=3
for i in `seq 1 $P`
do
cd /directory path/pool${i}
head -1 SummaryCounts.txt > SummaryCounts_sorted.txt
tail -n +2 SummaryCounts.txt | sed "s/Tag//g" | sort -k1,1n -k2,2n | awk
'BEGIN{OFS="\t";}{$1="Tag"$1;$2="Tag"$2; print $0;}' >> SummaryCounts_sorted.txt
cd ../
done
```

17. Count total number of sequences:

```
P=3
for i in `seq 1 $P`
do
cd /directory path/pool${i}
awk '{total = total + $4}END{print "Total sequences = "total}' /directory
path/pool${i}/SummaryCounts.txt
done
```

Example output:

Total sequences = 258816

Total sequences = 180541

Total sequences = 198568

18. Create a file of the summary counts split by PS info:

```
P=3
for pool in `seq 1 $P`
do
splitSummaryByPSInfo.py -p /directory path/PSInfofile.txt -l $pool -s
pool$pool/SummaryCounts_sorted.txt -o pool$pool/SummaryCounts_split.txt; done
```

Quality Filtering Sequences

19. Make a new directory to test specific filtering values, using the minimum number of PCR replicates accepted (e.g. 2 = filtering must remove sequences that occur in less than 2/3 PCR replicates) followed by the minimum number of reads accepted (e.g. 5 = filtering must remove sequences occurring in less than 5 replicates).

```
mkdir filter_minnumber_minnumber
```

20. Use the filter.py function to filter data for PCR replicatesand copy number:

filter.py -psInfo PSInfo filename.txt -x number of PCR replicates -y number of PCR reactions to accept -p number of pools -t 2 -l length of amplicon + tag + primer -o directory for specific filtering levels

This creates several output files: explain here

Check the files for mix ups, contents of extraction blanks, contents of positive control. Ideally blanks should be blank, positives should contain many reads.

21. Count the number of unique sequences to get an overall view of this information

```
awk '{h[$1]++}; END { for(k in h) print k, h[k] }'
Comparisons_2outOf3PCRs.countsThreshold2.txt
```

OTU Clustering

View the SequenceLengthDistribution.pdf to assess what the minimum and maximum length of sequences to retain should be. Choose a length cut-off based on this information.

22. Use the function in DAMe called convertToUSearch.py to…

```
module load convertToUSearch.py
convertToUSearch.py –h
```

```
convertToUSearch.py -i FilteredReads.fna -lmin minimum length -lmax maximum length
```

This command creates an output file called FilteredReads.for.sumaclust.fna

23. Use the programme Sumaclust to do OTU clustering

```
module load sumaclust/v1.0.20
```

```
sumaclust –h
```

```
sumaclust -t 0.96 FilteredReads.forsumaclust.fna > directory path/name of output.fna
```

Example output message:

```
Done : 100 %       234 clusters created.
```

This creates output files: explain here

24. Use the function tabulateSumaclust.py to convert the sumaclust output to a table form which can be used afterward by Blast.

module load tabulateSumaclust.py

tabulateSumaclust.py -h

tabulateSumaclust.py -s number to normalise -i file name.fna -o OutputFileName.txt -blast

This creates two output files, one which can be opened as a spreadsheet in e.g. excel (file ending .txt) and another which can be used to run a Blast (file ending in .txt.blast.txt).

Taxonomic assignment

25. Blast the OTU sequences using the .txt.blast.txt file against a blast databse on the UCPH server.

module load blast+/v2.6.0

blastn -query FileName.txt.blast.txt -out OutputFileName.output.txt -db nt –remote

26. Open MEGAN 6 community edition, and import the output file from the previous step. Select appropriate taxonomic rank, select all and export as cvs. Open created files and sort by OTU to merge with the OTU information.