THE UNIVERSITY OF HULL

Rapid biodiversity monitoring of freshwater ponds using environmental DNA: traversing the aquatic-terrestrial boundary in pondscapes

being a Thesis submitted for the degree of Doctor of Philosophy at the University of Hull

Lynsey Rebecca Harper

BSc (Hons), MSc, University of Glasgow

December 2018

Abstract

Environmental DNA (eDNA) analysis is transforming biodiversity monitoring in aquatic ecosystems with immense potential to inform their conservation and management. eDNA analysis is rapid, non-invasive, cost-efficient, and often more accurate and sensitive than conventional monitoring tools for single species detection and community survey. Ponds are extremely diverse yet understudied freshwater habitats that require novel tools to enable comprehensive, systematic, long-term monitoring. eDNA monitoring could radically improve assessments of pond biodiversity, but the applications and methodical constraints of this tool in ponds are largely unexplored. In this thesis, eDNA analysis was examined as a tool for monitoring biodiversity associated with ponds, including aquatic, semi-aquatic, and terrestrial taxa. eDNA analysis using metabarcoding was shown to have comparable detection sensitivity for *Triturus cristatus* to targeted eDNA analysis using quantitative PCR, depending on species detection thresholds applied. Using the community data generated by this method comparison, eDNA metabarcoding was validated as a tool for ecological hypothesis testing, specifically biotic and abiotic determinants of *T. cristatus* and vertebrate species richness. A novel eDNA assay was designed and validated for targeted survey of the threatened *Carassius carassius*, a fish species characteristic of ponds. Furthermore, eDNA metabarcoding was compared to established methods of freshwater invertebrate assessment, and all methods used to evaluate the impact of stocking *C. carassius* for conservation purposes. Finally, eDNA metabarcoding was vindicated as a tool to monitor semi-aquatic and terrestrial mammals associated with ponds, and investigate the spatiotemporal distribution of their eDNA signals in these water bodies as a function of behaviour. These results combined emphasise the biological and scientific importance of ponds, and the prospects of eDNA analysis - targeted and community approaches - for enhanced conservation, management, monitoring, and research of these valuable ecosystems.

Declaration of Authorship

I declare that the work herein is intellectually my own. Each data chapter received contributions from my supervisors (Bernd Hänfling, BH; Lori Lawson Handley, LLH) and collaborators (Christoph Hahn, CH; Neil Boonham, NB; Helen Rees, HCR; Kevin Gough, KCG; Erin Lewis, EL; Ian Adams, IPA; Peter Brotherton, PB; Susanna Phillips, SP; Nathan Griffiths, NPG; Carl Sayer, CDS; Daniel Read, DSR; Kirsten Harper, KJH; Rosetta Blackman, RCB; Jianlong Li, JL; Marco Benucci, MB; Matthew Hill, MJH; Angus Carpenter, AIC; Gill Murray-Dickson, GM-D; Cristina Di Muri, CDM; Callum Macgregor, CJM; Thomas Logan, TWL; Alan Law, AL; Thomas Breighthaupt, TB; Allan McDevitt, AM) as follows. All further assistance has been acknowledged as appropriate at the end of each chapter.

Chapters 2 & 3

I led the experiment, which was co-designed with BH, LLH, and NB. HCR, KCG, and NB contributed samples for processing. HCR performed extractions under licence from Natural England and quantitative PCR (qPCR) for environmental DNA (eDNA) samples from private contracts. NB performed qPCR for eDNA samples from Natural England's Great Crested Newt Evidence Enhancement Program. I performed remaining laboratory work and analysed the data. IPA and EL offered advice on and supervised sequencing. CH assisted with bioinformatics analysis. PB and SP contributed datasets for analysis. I wrote the first draft of each manuscript and was responsible for subsequent editing, under advice from supervisors and collaborators.

Chapter 4

I led the experiment, which was co-designed with BH, LLH, and CDS. CDS selected the study ponds in Norfolk and provided associated environmental data. I collected and filtered water samples with assistance from RCB, JL, and DSR. I designed the qPCR assay with assistance from KJH. I performed laboratory work with assistance from NPG and RCB. I analysed the data, wrote the first draft of the manuscript, and was responsible for subsequent editing, under advice from supervisors and collaborators.

Chapter 5

I led the experiment, which was co-designed with BH, LLH, and CDS. CDS selected the study ponds in Norfolk and provided associated environmental data. I built the custom reference sequence database for invertebrates with MB and RCB. I sampled the Norfolk ponds for invertebrates with assistance from RCB. I collected and filtered water samples from all study ponds with assistance from MB. DSR also assisted with filtration of water samples and provided advice on primers, PCR protocols, and sequencing. I performed laboratory work and analysed the data with advice from MJH. I wrote the first draft of the manuscript and was responsible for subsequent editing, under advice from supervisors and collaborators.

Chapter 6

I conceived the experiment, and designed it with BH and LLH. AIC and GM-D coordinated sampling at Wildwood Trust and Royal Zoological Society of Scotland Highland Wildlife Park respectively. I collected and filtered water samples with assistance from CDM, CJM, and TWL. AL, TWL, and TB helped select natural ponds to be surveyed using eDNA metabarcoding, camera trapping, and field signs, and provided camera traps for the experiment. I deployed camera traps with assistance from AL and TWL, following which I collected and analysed footage. I processed samples in the laboratory with advice from CDM and AM. DSR sequenced the final library. I completed bioinformatic processing of samples, and subsequent data analysis. I wrote the first draft of the manuscript and was responsible for subsequent editing, under advice from supervisors and collaborators.

Acknowledgements

This thesis was made possible by funding from the University of Hull. First and foremost, I would like to thank my supervisors, Bernd Hänfling and Lori Lawson Handley, for their continual guidance and support throughout this challenging and rewarding experience. I am grateful to you both for always making time to discuss ideas and problems, your encouragement when things got tough, and allowing me freedom and trust to develop as a scientist. It was a privilege to be your student.

I would like to thank my collaborators for their advice, support, and teaching me some of the skills required to complete my research. In particular, I count myself lucky to have worked with one of my environmental DNA idols, Helen Rees, who consistently gave encouragement and helped wherever possible. Daniel Read always went above and beyond, providing expert technical guidance and support. Carl Sayer was my pond guru - I will never again overlook a clump of trees situated in the middle of a farm field!

Many people I met at Hull were a source of strength and solace. My lab mates Marco Benucci, Graham Sellers, Cristina Di Muri, Rosetta Blackman, Jianlong Li, Alexander Austin, Nathan Griffiths, Hayley Watson, Harriet Johnson, and Peter Shum kept me sane and provided vital biscuits, cake, and hot chocolate as well as emotional and technical support. Current and former members of the EvoHull group, particularly Christoph Hahn, James Kitson, Africa Gómez, James Gilbert, Dave Lunt, Domino Joyce, Katharina Wollenberg-Valero, and Rob Donnelly, were an invaluable source of knowledge, critique, and discussion that improved my research.

I'm grateful to Sarah Argyle for her friendship and patient ear during our coffee and Zumba evenings. Pub Fridays with those already mentioned and Khia Dobbinson, Hannah West, Adam Bakewell, Lisa Malm, Rob Jacques, Ellen Moss and Emma Peasland provided some much needed laughter. Despite the distance between us, my Glaswegian friends continue to be there for me: Mairi Fraser, Kirsty Black, Rebbekah Dyvor, Amy Hendry, Laura Mills, Gordon Douglas, Luke Saunders, Miranda Chung, Christopher Stevenson, Nicola Skelly, Kirsty Wright, Finlay McGhee, Nikki Andrews, Mairi Hilton, Trudie Marshall, Steven Duncan, Andrew Nuttall, and Grant Walker. I especially want to

thank my Masters crew (PGDip!) for their unwavering friendship as several of us embark on scientific careers: Claire Harris, Tiffany Armstrong, Sarah Bierbaum-Williams, Christopher Pollock, Robert Paton, Iain Hill, Christopher Convery, Pascal Lovell, and Luis Enrique Hernandez Castro.

Most importantly, I would not have completed my PhD without my family. Thank you to my sister-in-law Tristan Higgins for your support and savvy perspective on what life throws at me. To my partner Callum Macgregor, one of the best things about coming to Hull was meeting you. You made this journey infinitely easier than it would have been alone. Thank you for always being my rock and making me smile with your shenanigans. To my sister Kirsten, thank you for blazing a trail in academia and environmental DNA before me. You taught me anything is possible if you keep trying and believe in yourself. I feel lucky to be in science with you and hope one day others will cite Harper & Harper. Finally, Mum and Dad: I could write another thesis solely about your influence on my life. The opportunities you gave me while growing up led me down the path I am on now. This path has not been without obstacles, but your unconditional love as well as emotional, financial, and practical support helped me overcome them all. Not many Dads would brave the "jaggies" and put up with antisocial hours to help their daughters look for newts! I could not ask or wish for better parents, and I am grateful for you always.

Contents

Abstract iii

Declaration of Authorship v

Acknowledgements vii

Contents ix

List of tables xvi

List of figures xix

Chapter 1: General Introduction 1

1.1 Freshwater biodiversity declines 3

1.2 The status of ponds 4

1.3 The value of ponds 5

1.4 Challenges to pond conservation and research 7

1.5 Environmental DNA (eDNA) analysis 7

1.6 Scope for pond eDNA monitoring and research 11

1.7 Thesis rationale 13

1.7.1 Can eDNA analysis be used to monitor threatened biodiversity associated with ponds? 14

1.7.2 Can eDNA metabarcoding be used to survey biodiversity at landscape level, including semi-aquatic and terrestrial taxa? 14

1.7.3 What are the prospects of eDNA metabarcoding for community investigation in ponds? 15

1.8 Data chapter summaries 15

1.8.1 Chapter 2: Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*) 16

1.8.2 Chapter 3: Ground-truthing environmental DNA (eDNA) metabarcoding for ecological hypothesis testing at the pondscape 16

1.8.3 Chapter 4: Development and application of environmental DNA surveillance for the threatened crucian carp (*Carassius carassius*) 17

1.8.4 Chapter 5: Assessing the impact of the threatened crucian carp (*Carassius carassius*) on pond invertebrate diversity - a comparison of conventional and molecular tools 17

1.8.5 Chapter 6: Environmental DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority mammals 18
Chapter 2: Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*) 19
Abstract 20
2.1 Introduction 21
2.2 Materials and methods 24
2.2.1 Sampling 24
2.2.2 Targeted qPCR for *T. cristatus* 24
2.2.3 Metabarcoding of vertebrate communities 25
2.2.4 Bioinformatic processing 26
2.2.5 Data analysis 26
2.2.5.1 Detection thresholds and contamination 26
2.2.5.2 Comparison of eDNA methods for *T. cristatus* detection 27
2.2.5.3 Cost and investigator effort 28
2.3 Results 29
2.3.1 Targeted qPCR and egg searches 29
2.3.2 Vertebrate metabarcoding 29
2.3.3 eDNA metabarcoding vs. qPCR for *T. cristatus* detection 30
2.3.4 Comparison of method cost and investigator effort 34
2.4 Discussion 37
2.4.1 Single-species detection by qPCR and metabarcoding 37
2.4.2 False negatives 40
2.4.3 False positives 40
2.4.4 Cost and investigator effort 42
2.4.5 Conclusion 42
2.5 Acknowledgements 43
2.6 Data accessibility 44
Chapter 3: Ground-truthing environmental DNA (eDNA) metabarcoding for ecological hypothesis testing at the pondscape 45
Abstract 46
3.1 Introduction 47
3.2 Materials and methods 49
3.2.1 Samples 49

3.2.2 DNA reference database construction 50
3.2.3 Primer validation 50
3.2.4 eDNA metabarcoding 50
3.2.5 Data analysis 51
3.3 Results 53
3.3.1 eDNA metabarcoding 53
3.3.2 Pondscape biodiversity 53
3.3.3 Biotic determinants of *T. cristatus* occurrence 54
3.3.4 Abiotic determinants of *T. cristatus* occurrence 59
3.3.5 Abiotic determinants of vertebrate species richness 59
3.3.6 Applicability of the HSI to *T. cristatus* and community eDNA surveys 63
3.4 Discussion 63
3.4.1 Pondscape biodiversity 63
3.4.2 Biotic determinants of *T. cristatus* occurrence 64
3.4.3 Abiotic determinants of *T. cristatus* occurrence 66
3.4.4 Abiotic determinants of vertebrate species richness 67
3.4.5 Applicability of the HSI to *T. cristatus* and community eDNA surveys 67
3.4.6 Limitations of eDNA metabarcoding 68
3.4.7 Prospects of eDNA metabarcoding for conservation, management, and research 69
3.5 Acknowledgements 69
3.6 Data accessibility 70
Chapter 4: Development and application of environmental DNA surveillance for the threatened crucian carp (*Carassius carassius*) 71
Abstract 72
4.1 Introduction 73
4.2 Materials and methods 76
4.2.1 Study sites 76
4.2.2. Conventional survey 77
4.2.3 eDNA sampling, capture and extraction 78
4.2.4 Assay design, specificity and sensitivity 79
4.2.5 Detection and quantification of *C. carassius* eDNA 81
4.2.6 DNA sequencing 82
4.2.7 Data analysis 83

4.3 Results 85
4.3.1 Assay specificity and sensitivity 85
4.3.2 qPCR analysis 88
4.3.3 Presence-absence detection 88
4.3.4 Factors influencing eDNA detection and quantification 89
4.4 Discussion 92
4.4.1 Using eDNA analysis for *C. carassius* conservation 93
4.4.2 Factors influencing eDNA detection and quantification 93
4.4.3 Optimisation of eDNA workflow 95
4.4.4 Concluding remarks 97
4.5 Acknowledgements 98
4.6 Data accessibility 98
Chapter 5: Assessing the impact of the threatened crucian carp (*Carassius carassius*) on pond invertebrate diversity – a comparison of conventional and molecular tools 99
Abstract 100
5.1 Introduction 101
5.2 Materials and methods 103
5.2.1 Study sites 103
5.2.2 Sweep-netting and morphotaxonomic identification 104
5.2.3 DNA metabarcoding samples 105
5.2.4 eDNA metabarcoding samples 106
5.2.5 Metabarcoding workflow 106
5.2.6 Data analysis 108
5.3 Results 112
5.3.1 Taxonomic composition by method 112
5.3.1.1 Sweep-netting and morphotaxonomic identification 112
5.3.1.2 DNA metabarcoding 112
5.3.1.3 eDNA metabarcoding 113
5.3.1.4 Combined methods 113
5.3.2 Impact of *C. carassius* stocking on pond invertebrates 114
5.3.3 Comparison of methods for freshwater invertebrate assessment 124
5.4 Discussion 126
5.4.1 Impact of *C. carassius* stocking on pond invertebrates 127
5.4.2 Comparison of methods for freshwater invertebrate assessment 129

5.4.3 Concluding remarks 134
5.5 Acknowledgements 135
5.6 Data accessibility 135
Chapter 6: Environmental DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority mammals 137
Abstract 138
6.1 Introduction 139
6.2 Materials and methods 142
6.2.1 Study species 142
6.2.2 Experiment 1: eDNA detection and signal strength in artificial systems 142
6.2.3 Experiment 2: eDNA detection and signal strength in natural systems 145
6.2.4 DNA extraction 147
6.2.5 eDNA metabarcoding 147
6.2.6 Data analysis 148
6.3 Results 152
6.3.1 eDNA metabarcoding 152
6.3.2 Experiment 1: eDNA detection and signal strength in artificial systems 152
6.3.3 Experiment 2: eDNA detection and signal strength in natural systems 156
6.4 Discussion 160
6.4.1 Experimental insights 161
6.4.2 Pitfalls of eDNA metabarcoding for mammal monitoring 163
6.4.3 Scope of eDNA metabarcoding for mammal monitoring 165
6.5 Acknowledgements 166
6.6 Data accessibility 167
Chapter 7: General Discussion 169
7.1. Can eDNA analysis be used to monitor threatened biodiversity associated with ponds? 171
7.2 Can eDNA metabarcoding be used to survey biodiversity at the pondscape, including semi-aquatic and terrestrial taxa? 175
7.3 What are the prospects of eDNA metabarcoding for community investigation in ponds? 179
7.4 Overcoming the limitations of eDNA metabarcoding 182
7.5 Future directions of eDNA monitoring in pond ecosystems 184
7.5.1 Biomonitoring 184

7.5.2 Population genetics and distinguishing hybrids 185
7.5.3 Disease management 186
7.5.4 Citizen science 186
7.6 Conclusions 187
References 189
Appendices 227
Appendix 2 229
Appendix 2.1: Supplementary methods 229
Appendix 2.2: Supplementary results 238
Appendix 2.3: Supplementary tables 240
Appendix 2.4: Supplementary figures 250
Appendix 2.5: Supplementary references 256
Appendix 3 259
Appendix 3.1: Supplementary methods 259
Appendix 3.2: Supplementary results 262
Appendix 3.3: Supplementary tables 263
Appendix 3.4: Supplementary figures 280
Appendix 3.5: Supplementary references 283
Appendix 4 284
Appendix 4.1: Supplementary methods 284
Appendix 4.2: Supplementary results and discussion 288
Appendix 4.3: Supplementary tables 289
Appendix 4.4: Supplementary figures 318
Appendix 4.5: Supplementary references 326
Appendix 5 328
Appendix 5.1: Supplementary methods 328
Appendix 5.2: Supplementary results 334
Appendix 5.3: Supplementary tables 335
Appendix 5.4: Supplementary figures 339
Appendix 5.5: Supplementary references 348
Appendix 6 350
Appendix 6.1: Supplementary methods 350
Appendix 6.2: Supplementary results 354
Appendix 6.3: Supplementary tables 355

Appendix 6.4: Supplementary figures..374
Appendix 6.5: Supplementary references 378

List of tables

Table 2.1: Summary of analyses testing for agreement between eDNA approaches 33
Table 2.2: Summary of analyses testing for variation in proportion of *T. cristatus* sequence reads 34
Table 3.1: Summary of established and novel abiotic and biotic determinants of *T. cristatus* occupancy 57
Table 3.2: Summary of analyses testing for variation in vertebrate species richness across ponds 61
Table 4.1: Top NCBI BLASTn hit for Sanger sequences 86
Table 4.2: Bayesian estimates of *C. carassius* eDNA occurrence probability at a pond (ψ), eDNA detection probability in a water sample (θ), and eDNA detection probability in a qPCR replicate (p) 90
Table 5.1: Summary of analyses (GLM) statistically comparing alpha diversity (taxon richness) at species-level and family-level between ponds 114
Table 5.2: Relative contribution of species turnover and nestedness to total beta diversity 117
Table 5.3: Summary of analyses (ANOVA) statistically comparing homogeneity of multivariate dispersions (MVDISP) between the communities in ponds 118
Table 5.4: Summary of analyses (PERMANOVA) statistically examining variation in community composition of ponds 122
Table 6.1: Summary of focal species studied at each wildlife park and their lifestyle. 143
Table S2.1: List of species for which no 12S rRNA records were available on GenBank 240
Table S2.2: List of species detected in PCR positive controls by eDNA metabarcoding 243
Table S2.3: Summary statistics for each Illumina MiSeq run 244
Table S2.4: Summary of read counts 244
Table S2.5: Summary of species detected by eDNA metabarcoding 245
Table S2.6: Summary of contaminants detected in PCR negative, or No Template Controls (NTCs) 248
Table S2.7: Summary of agreement (+) and disagreement (-) between egg searches, qPCR NT, qPCR TA, metabarcoding NT, and metabarcoding TA for *T. cristatus* detection 249

Table S3.1: Summary of environmental metadata on pond characteristics and surrounding terrestrial habitat........263
Table S3.2: List of species for which no 12S rRNA records were available on GenBank 265
Table S3.3: List of species detected in PCR positive controls by eDNA metabarcoding 267
Table S3.4: Effect of number of species in different vertebrate groups on *T. cristatus* occupancy........268
Table S3.5: Summary of different significant associations between *T. cristatus* and other vertebrate species 270
Table S3.6: Summary of abiotic and biotic determinants of *T. cristatus* occupancy....272
Table S3.7: Summary of relationship between HSI score and *T. cristatus* occupancy 274
Table S3.8: Summary of abiotic determinants of vertebrate species richness 275
Table S3.9: Summary of relationship between HSI score and vertebrate species richness 277
Table S3.10: Summary of species detected by eDNA metabarcoding 278
Table S4.1: Summary of eDNA analysis for each sample collected from ponds in Norfolk 289
Table S4.2: Corresponding species, accession number, and geographic location for sequences that were downloaded from the NCBI nucleotide database........293
Table S4.3: List of non-target species tested using PCR and qPCR to validate assay specificity for *C. carassius*........314
Table S4.4: Summary of eDNA amplification by PCR and qPCR........315
Table S4.5: Summary of model-selection criteria (PPLC and WAIC) for each model 316
Table S5.1: Summary of number of invertebrate species detected by each method across each invertebrate group 335
Table S5.2: Summary of number of invertebrate families detected by each method across each invertebrate group 337
Table S6.1: Behavioural observation data for species housed at wildlife parks........355
Table S6.2: Summary of directed, random, or other samples collected for each species at wildlife parks........359
Table S6.3: Summary of exotic amphibian species housed in the wet laboratory at the University of Kent........361

Table S6.4: Summary of samples collected from natural ponds at locations where target species were confirmed as present 362
Table S6.5: List of taxa detected in PCR positive controls by eDNA metabarcoding . 368
Table S6.6: Summary of species detected using eDNA metabarcoding across all samples collected in this study 369
Table S6.7: Summary of detection rates for species which were detected by at least one survey method 372

List of figures

Figure 1.1: Schematic of eDNA workflow for samples collected from ponds 10
Figure 2.1: Adult male great crested newt (*Triturus cristatus*) 23
Figure 2.2: Comparison of survey methodology for *T. cristatus* detection 31
Figure 2.3: Venn diagram which summarises the number of positive *T. cristatus* detections 32
Figure 2.4: Relationship between fixed effects (qPCR score, post-PCR eDNA concentration) and response variable (proportion of *T. cristatus* reads) in eDNA samples 35
Figure 2.5: Cost and investigator effort required for targeted qPCR of *T. cristatus* and eDNA metabarcoding of vertebrate communities 36
Figure 3.1: *T. cristatus* presence (orange) and absence (grey) in relation to number of species from different vertebrate groups detected by eDNA 55
Figure 3.2: Heat map showing significant ($P < 0.05$) positive and negative species associations 56
Figure 3.3: Biotic and abiotic determinants of *T. cristatus* occurrence 60
Figure 3.4: Abiotic determinants of vertebrate species richness 62
Figure 4.1: A crucian carp (*Carassius carassius*) individual (a) and examples of the study ponds (b-d) 74
Figure 4.2: Map of pond locations in North Norfolk, eastern England 77
Figure 4.3: Alignment of consensus sequences for a region of the mitochondrial cytochrome *b* (*cytb*) gene in 24 European freshwater fishes, including *C. carassius* 80
Figure 4.4: Estimated probability of eDNA detection in qPCR replicates 91
Figure 4.5: Relationship between fixed effects and response variable (DNA copy number) in ponds 92
Figure 5.1: Mean alpha diversity (taxon richness) of invertebrates in ponds 115
Figure 5.2: Mean alpha diversity (taxon richness) at species-level (a) and family-level (b) of the different invertebrate groups 116
Figure 5.3: Non-metric Multidimensional Scaling (NMDS) plots of species-level invertebrate communities 120
Figure 5.4: Non-metric Multidimensional Scaling (NMDS) plots of family-level invertebrate communities 121

Figure 5.5: The relative contribution of biotic (*C. carassius* presence-absence) and abiotic variables (pond area) to species-level turnover (a) and total beta diversity (b) 123
Figure 5.6: Venn diagram which summarises the number of invertebrate species (a) and families (b) detected across the 18 study ponds 124
Figure 5.7: Mean alpha diversity (taxon richness) of invertebrates in ponds across Norfolk and East Yorkshire 125
Figure 5.8: Non-metric Multidimensional Scaling (NMDS) plots of invertebrate communities 126
Figure 6.1: Heatmap showing proportional read counts for samples collected from focal species enclosures at wildlife parks 153
Figure 6.2: Relationships predicted by the binomial GLMMs between proportional read counts and sample type nested within species (a) or species lifestyle (b) 154
Figure 6.3: Boxplots showing the mean proportional read counts for specific behaviours exhibited by different focal species (a) and behaviour type (b) 155
Figure 6.4: Camera trap photographs taken at natural ponds where focal species were confirmed as present 156
Figure 6.5: Tile plot showing species presence-absence at natural ponds as indicated by field signs, camera trapping, and eDNA metabarcoding 157
Figure 6.6: Heatmap showing proportional read counts for samples collected from natural ponds at sites where focal species presence was confirmed 159
Figure 6.7: Heatmap showing species detected from samples collected at ponds (THL01 and THL02) within Tophill Low Nature Reserve every 24 hrs over a 5-day period (D01 - D05) 160
Figure S2.1: Gel image showing results of *in vitro* primer validation for each species 250
Figure S2.2: Heat maps of sequence read distribution for taxonomic assignments in each vertebrate group across all eDNA samples 251
Figure S2.3: Proportion of eDNA samples in which each species was detected by eDNA metabarcoding. 252
Figure S2.4: Presence of foreign DNA in PCR negative controls across sequencing runs 253
Figure S2.5: Presence of cichlid (*Rhamphochromis esox*) DNA (PCR positive control) amongst PCR negative controls and eDNA samples 254
Figure S2.6: Presence of human (*Homo sapiens*) DNA amongst PCR controls and eDNA samples 255

Figure S3.1: Location of ponds (*n* = 504) sampled for eDNA as part of Natural England's Great Crested Newt Evidence Enhancement Programme 280
Figure S3.2: Gel image showing results of *in vitro* primer validation 281
Figure S3.3: Occurrence of *T. cristatus* in relation to species from different vertebrate groups 282
Figure S4.1: PCR products resulting from annealing temperature gradient PCR performed to test specificity of primers for crucian carp (*C. carassius*) against non-target species 318
Figure S4.2: PCR products for eDNA samples from ponds recently stocked with crucian carp (*C. carassius*) in Norfolk 319
Figure S4.3: qPCR amplification plot for test of primer and probe specificity for crucian carp (*C. carassius*) against 10 non-target fish species 320
Figure S4.4: qPCR standard curve plot for test of assay sensitivity 321
Figure S4.5: Alignment of good quality Sanger sequences obtained for qPCR amplicons 322
Figure S4.6: Variation amongst eDNA samples for each pond 323
Figure S4.7: Relationship between fixed effects and response variable (DNA copy number) in ponds 324
Figure S4.8: PCR products of gBlocks® standards and five eDNA samples from two ponds 325
Figure S5.1: Barplot summarising the number and proportion of species with and without reference sequences from GenBank for each custom invertebrate database 339
Figure S5.2: Barplot summarising the number of species with and without records on GenBank according to invertebrate groups 340
Figure S5.3: Gel images showing results of *in vitro* primer validation for primers mICOIintF and jgHCO2198 342
Figure S5.4: Gel images showing results of PCR for primers BF2 and BR2 343
Figure S5.5: Gel images showing results of PCR for primers fwhF1 and fwhR1 344
Figure S5.6: Gel image showing results of annealing temperature gradient PCR for primers mICOIintF and jgHCO2198 345
Figure S5.7: Gel images showing results of PCR cycle number optimisation for primers mICOIintF and jgHCO2198 346
Figure S5.8: Venn diagrams which summarise the number of species (a) and families (b) detected within the major invertebrate groups across the 18 study ponds 347

Figure S6.1: Example gel image of pooled first PCR products 374
Figure S6.2: Heatmap showing the frequency of contamination in negative process controls 375
Figure S6.3: Barplot showing the impact of different false positive sequence thresholds on the proportion of taxa detected in each sample 376
Figure S6.4: Relationship between the fixed effects (volume and number of filters) and response variable (proportional read count) 377

Chapter 1: General Introduction

Pond ecosystem in North Norfolk, eastern England

Certain content in sections 1.2 and 1.4 – 1.6 of this chapter was written for a review paper on eDNA monitoring in ponds that was first-authored by LRH and published in *Hydrobiologia* as

Harper, L.R., Buxton, A.S., Rees, H.C., Bruce, K., Brys, R., Halfmaerten, D., Read, D.S., Watson, H.V., Sayer, C.D., Jones, E.P., Priestley, V., Mächler, E., Múrria, C., Garcés-Pastor, S., Medupin, C., Burgess, K., Benson, G., Boonham, N., Griffiths, R.A., Lawson Handley, L. & Hänfling, B. (2019) Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. *Hydrobiologia*, **826**, 25–41.

1.1 Freshwater biodiversity declines

Freshwater ecosystems are hotspots of biodiversity (Strayer & Dudgeon, 2010), containing 6-10% of global biodiversity (~125,000 species) whilst occupying less than 1% of the earth's surface (Dudgeon *et al.*, 2006; Abell *et al.*, 2008; Balian *et al.*, 2008; Pittock, Hansen & Abell, 2008; Strayer & Dudgeon, 2010; Geist, 2011; Collen *et al.*, 2014; Mantyka-Pringle *et al.*, 2014). In the UK, freshwater habitats support over 50% of native aquatic taxa and provide habitat for many rare species (Sayer *et al.*, 2012), including some on the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (Céréghino *et al.*, 2008). However, freshwater organisms are experiencing a greater rate of decline than marine or terrestrial organisms (Sala *et al.*, 2000; Dudgeon *et al.*, 2006; Pittock *et al.*, 2008; Williams, Whitfield & Biggs, 2008; Strayer & Dudgeon, 2010; Geist, 2011), with one third of freshwater species at higher risk of extinction than terrestrial counterparts (Collen *et al.*, 2014). Indeed, 25% of freshwater species are classed as threatened or regionally extinct with changing distribution and abundance (Mächler *et al.*, 2014).

Freshwater ecosystems have not been studied to the same extent as terrestrial ecosystems and existing data is biased in terms of geographic, habitat, and taxonomic coverage, thus losses may be higher than current estimates (Dudgeon *et al.*, 2006; Strayer & Dudgeon, 2010). There is little time to rectify this bias as widespread, irreversible environmental degradation has already been incurred by anthropogenic activity, including globalisation, climate change, human movement and expansion, spread of invasive non-native species, and resource exploitation (Brautigam, 1999; Sala *et al.*, 2000; Brönmark & Hansson, 2002; Dudgeon *et al.*, 2006; Butchart *et al.*, 2010). In the absence of mitigation, this degradation is only likely to increase as the human population continues to grow (Strayer & Dudgeon, 2010). Climate change processes will also accelerate rates of extinction and freshwater habitat loss (Pittock *et al.*, 2008; Oertli *et al.*, 2009; Strayer & Dudgeon, 2010; Geist, 2011). Therefore, coordinated efforts using integrative strategies at catchment-scale are needed for effective conservation, management, monitoring, and rehabilitation of extant freshwater biodiversity (Dudgeon *et al.*, 2006).

1.2 The status of ponds

Globally, there are an estimated 64 million to 3 billion ponds or small lakes (Downing *et al.*, 2006; McDonald *et al.*, 2012; Bartout *et al.*, 2015; Biggs, von Fumetti & Kelly-Quinn, 2016; Hill *et al.*, 2018), with ponds outnumbering larger lentic ecosystems approximately 100:1 (Downing *et al.*, 2006; Céréghino *et al.*, 2008). Ponds represent a high proportion of global freshwater habitat despite their limited size, comprising up to 30% of standing freshwater by area (Downing *et al.*, 2006). In the UK alone, an estimated 800,000 ponds comprise 97% of standing water bodies and 14% of total surface water area (Wood, Greenwood & Agnew, 2003). These small water bodies occur in all land-use types at high frequency (Céréghino *et al.*, 2008) and possess ecological, aesthetic, and recreational value (Gee *et al.*, 1997; Wood *et al.*, 2003; Goertzen & Suhling, 2012; Biggs *et al.*, 2016). Ponds can be natural or manmade (Gee *et al.*, 1997; Céréghino *et al.*, 2008; Davies *et al.*, 2008) as well as permanent or temporary (Céréghino *et al.*, 2008).

Paradoxically, small size and high occurrence led to many ponds being drained, and accelerated pond loss over several decades (Brönmark & Hansson, 2002; Boix *et al.*, 2012). For instance, Beebee (1997) documented a decrease of 21% in chalk Downs dewpond numbers in southern England between 1977 and 1996. Similarly, Heath & Whitehead (1992) estimated 55-69% of ponds were lost in Essex from 1870-1989. This was echoed by Boothby & Hull (1997) who observed a considerable drop (61%) in Cheshire pond numbers between 1870 and the 1990's. Declines are primarily due to land use intensification and development, infilling for agriculture, agricultural run-off, pollution, stocking of fish or wildfowl, and invasion by non-native plant species, which ponds are ill-equipped to handle due to their small size and restricted buffering capacity (Biggs *et al.*, 1996, 2016; Brönmark & Hansson, 2002; Williams *et al.*, 2010). However, the creation of new ponds may have counteracted declines in the UK as an annual net increase (1.4%) in pond numbers was recorded between 1998 and 2007 (Williams *et al.*, 2010).

Ponds may be isolated from one another or form linked networks within the landscape. Until recently, pondscapes – a pond, its immediate catchment, and the terrestrial matrix of land between ponds – were poorly understood (Boothby, 1997; Wood *et al.*, 2003; Hill *et al.*, 2018). Ponds were not mentioned or included in the European Water Framework Directive 2000/60/EC (European Commission, 2000; Davies *et al.*, 2008; Hill *et al.*, 2018), and have been neglected in research, scientific monitoring, and

policy (Biggs *et al.*, 2005; De Meester *et al.*, 2005; Céréghino *et al.*, 2008; Oertli *et al.*, 2009; Boix *et al.*, 2012; Hill *et al.*, 2018), despite being threatened by anthropogenic activity and environmental change, and having greater vulnerability to environmental stressors than larger water bodies with larger catchments (Biggs *et al.*, 2016). This has been somewhat rectified with the establishment of the European Pond Conservation Network (EPCN) in 2004, but pond research has continued to lag behind that of other freshwater ecosystems (Oertli *et al.*, 2009; Boix *et al.*, 2012). Long-term monitoring of pond networks is particularly rare (Oertli *et al.*, 2009), although the designation of ponds as a "Priority Habitat" in the UK may increase incentive for their routine monitoring here (Joint Nature Conservation Committee & Defra, 2012; Hill *et al.*, 2018).

1.3 The value of ponds

Ponds are hugely valuable in terms of biodiversity: these systems are integral to invertebrate, plant, and amphibian diversity, and form a key component of the terrestrial habitat matrix (Wood *et al.*, 2003; Hassall, Hollinshead & Hull, 2012). Ponds act as stepping stones for a wide variety of aquatic and terrestrial taxa to larger water bodies and enable dispersal across landscapes (Oertli *et al.*, 2009; Hassall *et al.*, 2012) by providing opportunities for drinking, foraging, and reproduction (Almeida *et al.*, 2013; Biggs *et al.*, 2016; Klymus *et al.*, 2017b). Consequently, these water bodies provide critical habitat for biodiversity in a fragmented landscape (Céréghino *et al.*, 2008) and support many rare, protected, and unique species (Wood *et al.*, 2003; Hill *et al.*, 2018), including 80 UK Biodiversity Action Plan (BAP) species (Williams *et al.*, 2010).

In arable lowlands of the UK, ponds supported more plant and invertebrate species than rivers, streams, and ditches, as well as more uncommon species and distinct invertebrate assemblages. This was believed to result from the vast physicochemical heterogeneity and greater degree of isolation that ponds experience (Williams *et al.*, 2003). These patterns were reaffirmed at European level, where individual ponds supported the highest number of macrophyte and macroinvertebrate species and made the greatest contribution to regional species richness. Ponds supported more unique species and subsequently possessed high beta (between-site) diversity. Ponds also had greater gamma (landscape) diversity in spite of high alpha (site) diversity observed in rivers (Davies *et al.*, 2008). Undoubtedly, the biodiversity these ecosystems support must be

examined in the context of individual ponds and pond networks, but can only be maintained if influencers and stressors of these systems are understood (Wood *et al.*, 2003; De Meester *et al.*, 2005; Céréghino *et al.*, 2008; Oertli *et al.*, 2009; Boix *et al.*, 2012; Biggs *et al.*, 2016).

Ponds not only have tremendous biodiversity value, but also enormous scientific value as small and abundant ecosystems along broad ecological gradients, enabling experimental validation and hypothesis testing in ecology and conservation (De Meester *et al.*, 2005). Many ponds are threatened by anthropogenic activity to different degrees and reflect changes in ecosystem health. Consequently, these small water bodies can act as early warning systems for long-term effects in larger water bodies, e.g. lakes (De Meester *et al.*, 2005; Oertli *et al.*, 2009). Ponds also function as aquatic islands in a terrestrial landscape of anthropogenic activity, and represent patches of good quality habitat in an inhabitable matrix (De Meester *et al.*, 2005; Céréghino *et al.*, 2008). This is the foundation of metapopulation, metacommunity, and metaecosystem theory. Individuals of different species move between ponds (metapopulations), and the communities of different ponds are linked by multi-species dispersal and interaction (metacommunities). These ecosystems are thereby connected by the spatial exchange of resources and organisms, i.e. metaecosystems (Gounand *et al.*, 2018). Consequently, ponds are model systems for studies of landscape characteristics and connectivity (De Meester *et al.*, 2005; Céréghino *et al.*, 2008).

Beyond their ecological advantages, ponds have logistical advantages for inclusion in scientific research. These small water bodies can be sampled with ease repeatedly and quantitatively, and samples are more representative than those taken from larger water bodies (De Meester *et al.*, 2005; Céréghino *et al.*, 2008). In comparison to larger lakes, ponds exhibit less spatial heterogeneity and experience less weather interference. Surveys can be standardised with relative ease compared to large water bodies, but inter-year variability may be higher in ponds, e.g. hydroperiod (De Meester *et al.*, 2005). Ponds can be simulated in mesocosms or containers (De Meester *et al.*, 2005) as well as artificially created with relative ease (Williams *et al.*, 2008). These model systems create opportunities for replicated experiments under controlled environmental conditions, allowing for more complex experimental design and hypothesis testing (De Meester *et al.*, 2005).

1.4 Challenges to pond conservation and research

Exhaustive assessment and systematic monitoring of pond biodiversity has been hindered by the cost, time, and taxonomic expertise required to survey these abundant water bodies (Briers & Biggs, 2005; Hill *et al*., 2018). Often data is at genus- or family-level when species-level knowledge is required for effective conservation. As a result, indicator taxa (e.g. plants, water beetles, molluscs, dragonflies, amphibians, fish) have been selectively monitored to assess ecosystem health. However, these taxa do not always reflect the response or trends of the wider biological community (Gustafson, Pettersson & Malmgren, 2006; Sewell & Griffiths, 2009; Goertzen & Suhling, 2012; Thomsen *et al.*, 2012; Evans *et al.*, 2016b). Large-scale community-level monitoring, encompassing alpha, beta, and gamma diversity analyses, would provide more comprehensive understanding of biodiversity in changing environments (Hajibabaei *et al.*, 2016). Standardised, sensitive sampling methods that can detect rare and low-density species, maximise taxon richness, and minimise sampling effort (Céréghino *et al.*, 2008; Oertli *et al.*, 2009; Hajibabaei *et al.*, 2016), are required to ensure rapid, accurate and contemporary records of ecosystem biodiversity, health, and function at all scales (Baird & Hajibabaei, 2012; Hajibabaei *et al.*, 2016). In this context, molecular tools offer a solution through rapid, sensitive, cost-effective, non-invasive monitoring, and promise to enhance our understanding of global biodiversity (Hajibabaei *et al.*, 2016). One tool in particular is at the frontier of aquatic biodiversity monitoring: environmental DNA (eDNA) analysis.

1.5 Environmental DNA (eDNA) analysis

Environmental DNA (eDNA) analysis is the identification of species using the genetic material that organisms deposit in their environment (Valentini, Pompanon & Taberlet, 2009; Bohmann *et al.*, 2014; Rees *et al.*, 2014b; Thomsen & Willerslev, 2015). In aquatic ecosystems, DNA can be released into the water column via skin cells, saliva, hair, mucus, blood, urine, faeces, gametes, or decomposition, and diffuses rapidly thereby increasing detection probability (Valentini *et al.*, 2009; Bohmann *et al.*, 2014; Rees *et al.*, 2014b; Thomsen & Willerslev, 2015). Mobile organisms can also transfer their DNA to water bodies, thus eDNA is representative of species in the surrounding area (Rees *et al.*,

2014b). eDNA persistence is highly variable depending on the environment itself, e.g. weeks in water, decades in sediment, or thousands of years in permafrost (Thomsen & Willerslev, 2015). eDNA is more contemporary of species presence in water due to degradative processes accelerated by water chemistry (Strickler, Fremier & Goldberg, 2015; Goldberg, Strickler & Fremier, 2018; Seymour *et al.*, 2018), temperature (Takahara *et al.*, 2012; Strickler *et al.*, 2015; Eichmiller, Best & Sorensen, 2016; Robson *et al.*, 2016; Buxton *et al.*, 2017b), exposure to ultraviolet (UV) light (Strickler *et al.*, 2015), trophic state (Klymus *et al.*, 2015; Eichmiller *et al.*, 2016) and microbial activity (Barnes *et al.*, 2014; Tsuji *et al.*, 2017; Salter, 2018).

eDNA analysis has provided ecologists with unprecedented power to detect single species or describe whole communities (Lawson Handley, 2015). Typically, DNA is extracted from environmental samples (e.g. water, soil, air) and short DNA fragments (<500 bp) are amplified using Polymerase Chain Reaction (PCR), following which amplified products are sequenced to determine species identity (Thomsen & Willerslev, 2015). Single species can be targeted with species-specific primers using PCR, real-time quantitative PCR (qPCR), or droplet digital PCR (ddPCR), whereas entire communities can be passively monitored with conserved primers using PCR and high-throughput sequencing, i.e. eDNA metabarcoding (Fig. 1; Valentini *et al.*, 2009; Taberlet *et al.*, 2012; Bohmann *et al.*, 2014; Rees *et al.*, 2014b; Lawson Handley, 2015; Thomsen & Willerslev, 2015; Deiner *et al.*, 2017). Ponds were the first natural habitats screened for macro-organism eDNA by Ficetola *et al.* (2008), who demonstrated reliable detection of the invasive American bullfrog (*Lithobates catesbeianus*), even at low densities. Since this initial publication, a large and growing number of studies have utilised eDNA in a range of environments (reviewed for example by Rees *et al.*, 2014b; Lawson Handley, 2015; Thomsen & Willerslev, 2015; Deiner *et al.*, 2017).

eDNA approaches are often more affordable and logistically feasible than conventional counterparts (Biggs *et al.*, 2014; Davy, Kidd & Wilson, 2015; Sigsgaard *et al.*, 2015; Smart *et al.*, 2016; Evans *et al.*, 2017b), and have enormous potential to enable ecological study at greater temporal and spatial scales (Jerde *et al.*, 2011; Biggs *et al.*, 2015; Deiner *et al.*, 2016; Kelly *et al.*, 2016; Bista *et al.*, 2017; Minamoto *et al.*, 2017; Bálint *et al.*, 2018; Grey *et al.*, 2018; Nakagawa *et al.*, 2018). They are non-invasive and minimise risk of spreading disease or invasive non-native species (Tréguier *et al.*, 2014; Sigsgaard *et al.*, 2015; Valentini *et al.*, 2016). eDNA sampling is independent of weather conditions and could enable year-round monitoring by detecting organisms during

periods of low density or at different developmental stages (Rees *et al.*, 2014a, 2017; Buxton, Groombridge & Griffiths, 2018). Consequently, eDNA analysis has been heralded as a new tool for conservation and management purposes (Deiner *et al.*, 2017) and may resolve problems encountered with conventional monitoring tools in complex and species-rich systems (Lopes *et al.*, 2016; Ishige *et al.*, 2017; Sasso *et al.*, 2017; Bálint *et al.*, 2018). This tool is increasingly used for survey and detection of aquatic vertebrates (see reviews by Rees *et al*., 2014b; Lawson Handley, 2015; Thomsen & Willerslev, 2015; Deiner *et al*., 2017), but could also ensure aquatic invertebrate monitoring at high resolution, standardised by molecular protocols, and independent of specimen collection and taxonomic expertise (Mächler *et al.*, 2014; Deiner *et al.*, 2016; Bista *et al.*, 2017; Blackman *et al.*, 2017; Elbrecht *et al.*, 2017b; Macher & Leese, 2017; Niemiller *et al.*, 2017; Macher *et al.*, 2018).

eDNA metabarcoding in particular can provide rich, reproducible, and spatially consistent biodiversity data (Deiner *et al.*, 2016; Port *et al.*, 2016; Cilleros *et al.*, 2018), and continues to gain traction in freshwater monitoring with studies on lotic and lentic ecosystems (Civade *et al.*, 2016; Hänfling *et al.*, 2016; Lopes *et al.*, 2016; Olds *et al.*, 2016; Shaw *et al.*, 2016a; Vences *et al.*, 2016; Craine *et al.*, 2017; Sasso *et al.*, 2017; Cilleros *et al.*, 2018; Li *et al.*, 2018a; Nakagawa *et al.*, 2018). eDNA metabarcoding has been successfully used in ponds to survey temperate and tropical amphibian communities (Valentini *et al.*, 2016; Bálint *et al.*, 2018), fish assemblages (Valentini *et al.*, 2016; Evans *et al.*, 2017a; Li *et al.*, 2018b, c), and has strong capacity to detect semi-aquatic and terrestrial species (Chapters 3 & 6; Ushio *et al.*, 2017, 2018; Klymus *et al.*, 2017b; Harper *et al.*, 2018b). In contrast to vertebrates, published work on eDNA metabarcoding of pond invertebrates is distinctly lacking despite strong interest in this sector. A handful of studies successfully detected a range of macroinvertebrate taxa from running water (Deiner *et al.*, 2016; Blackman *et al.*, 2017; Klymus, Marshall & Stepien, 2017a; Macher & Leese, 2017; Macher *et al.*, 2018) and lakes (Bista *et al.*, 2015, 2017), but these taxa often comprise a low proportion of total sequence reads if conserved primers that amplify diverse taxonomic groups are used (Deiner *et al.*, 2016; Macher & Leese, 2017; Macher *et al.*, 2018). Metabarcoding has yet to be routinely implemented for pond surveys, but has a number of applications which could improve our knowledge and understanding of pond biodiversity.

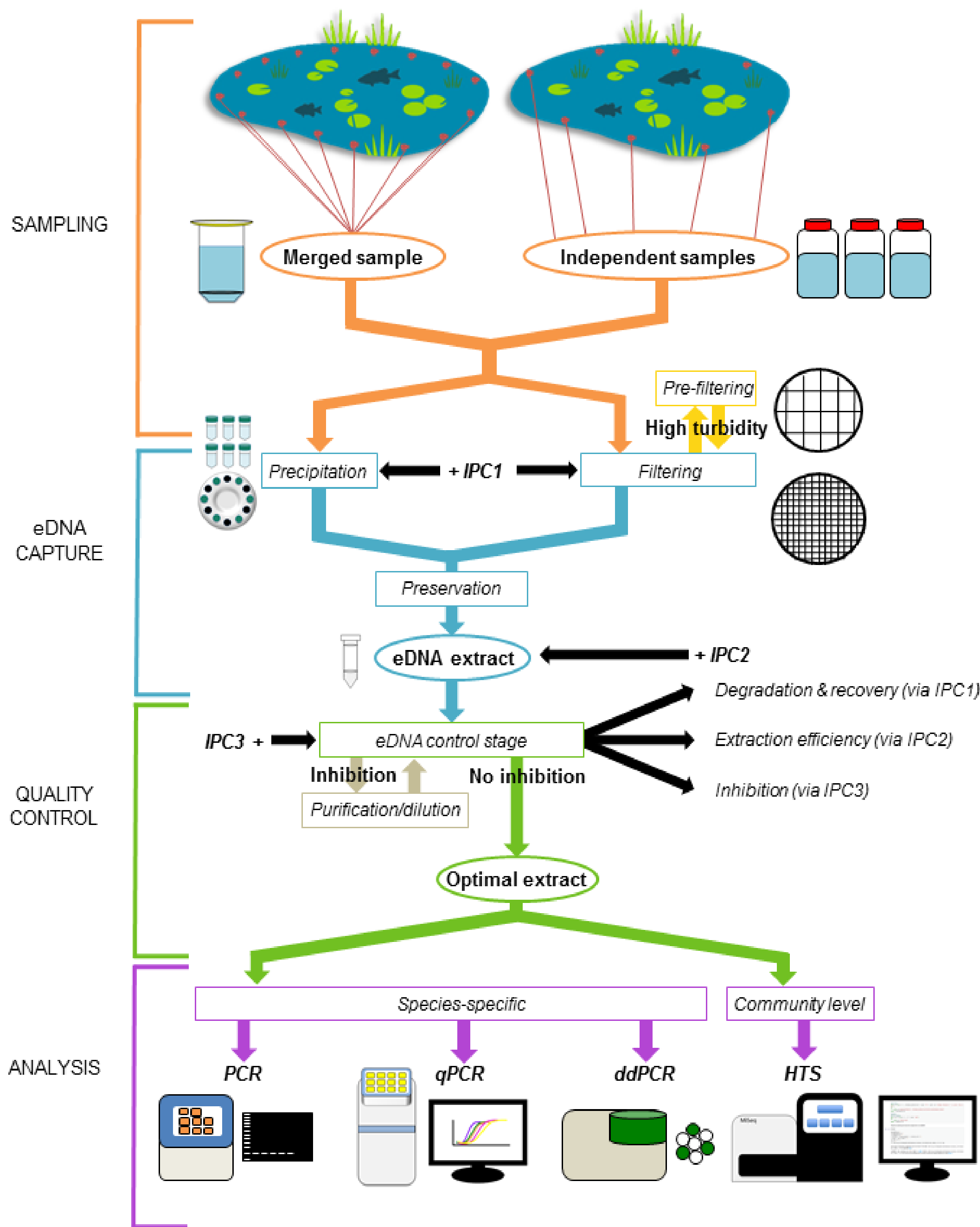


Figure 1.1: Schematic of eDNA workflow for samples collected from ponds. Three different Internal Positive Controls (IPCs) are recommended for inclusion during the stages of eDNA capture and quality control to identify substandard samples which require reanalysis or resampling. Pre-filtering is recommended if water samples are turbid. Figure reproduced with permission from Harper *et al*. (2019a).

1.6 Scope for pond eDNA monitoring and research

We are only beginning to realise the potential of eDNA analysis for pond monitoring and research. Undoubtedly, eDNA analysis could enhance biological recording and assessment of pond biodiversity. This molecular tool can complement or outperform conventional methods of monitoring pond biodiversity (Thomsen *et al.*, 2012; Takahara, Minamoto & Doi, 2013; Biggs *et al.*, 2015; Valentini *et al.*, 2016; Bálint *et al.*, 2018; Kuzmina, Braukmann & Zakharov, 2018; Mauvisseau *et al.*, 2018). Indeed, the work of Thomsen *et al*. (2012) on ponds and other freshwater habitats was pivotal to the development of targeted eDNA surveillance for many rare and endangered species across the globe (Bellemain *et al.*, 2016; Simpfendorfer *et al.*, 2016; Bylemans *et al.*, 2017; Doi *et al.*, 2017; Niemiller *et al.*, 2017; Torresdal, Farrell & Goldberg, 2017; Weltz *et al.*, 2017; Hunter *et al.*, 2018). Targeted eDNA analysis has shown enormous potential for distribution mapping as well as relative abundance and biomass estimation of amphibians and fish in ponds (Takahara *et al.*, 2012; Thomsen *et al.*, 2012; Biggs *et al.*, 2015; Buxton *et al.*, 2017b).

In the UK, eDNA analysis of pond water was first implemented for the great crested newt (*Triturus cristatus*, Rees *et al.*, 2014a). Following this initial work, a national citizen science monitoring scheme implementing eDNA analysis was launched for *T. cristatus*. This work evidenced that eDNA analysis can deepen our understanding of species distribution patterns and activity, where large-scale eDNA sampling informed distribution modelling for *T. cristatus* (Biggs *et al.*, 2015). Thereafter, eDNA analysis was formally recognised as a survey tool for this legally protected species (Natural England, 2015), and eDNA survey results now underpin new Natural England strategic licensing policies that aim to provide landscape-level species protection for *T. cristatus* (Harper *et al.*, 2019a). *T. cristatus* is an excellent example of eDNA monitoring in practice, and has contributed to the adoption of eDNA analysis for targeted survey of other pond biota (Davy *et al.*, 2015; Fujiwara *et al.*, 2016; Matsuhashi *et al.*, 2016; Newton *et al.*, 2016; Robson *et al.*, 2016; Geerts *et al.*, 2018; Goldberg *et al.*, 2018; Mauvisseau *et al.*, 2018; Raemy & Ursenbacher, 2018; Harper *et al.*, 2019b).

Ponds are often considered to be closed systems, but may receive inputs from inflow, land surface run-off (especially during high rainfall and flood events), and mobile species (e.g. birds, dragonflies, amphibians, water beetles). They can therefore act as natural samples of biodiversity in the wider environment and provide information on

entire ecosystems (De Meester *et al.*, 2005). Ponds are also good indicators of the quality of their local environment, being impacted both directly and indirectly (through large aquatic-terrestrial contact zones) by anthropogenic and environmental stressors (De Meester *et al.*, 2005). Consequently, eDNA metabarcoding of pond water can reveal the impact of multiple stressors on a broad range of taxa. For example, eDNA metabarcoding revealed wildlife using uranium mine containment ponds as water sources, and supplemented conventional assessment of ecotoxicological effects of uranium mining on local biodiversity (Klymus *et al.*, 2017b).

Beyond a step change in biodiversity monitoring and research, eDNA analysis in ponds offers endless experimental opportunities to heighten understanding of eDNA dynamics due to the vast physical and chemical heterogeneity of these ecosystems. Pond water is comparatively stagnant, and the lack of flow and relatively small water volumes in ponds allows eDNA to accumulate over time to concentrations not attainable in most other water bodies. This has benefits for the amount of target DNA present and subsequent detection probability (Buxton, Groombridge & Griffiths, 2017a). However, eDNA accumulation can reduce ability to distinguish contemporary from recent or historic presence (Rees *et al.*, 2014b). Under stagnant conditions eDNA can settle out of suspension, but once again become incorporated into the water column following sediment disturbance (Turner, Uy & Everhart, 2015; Buxton *et al.*, 2018). eDNA may remain detectable in ponds for several weeks under 'optimal' conditions (Buxton *et al.*, 2017a), but can also degrade rapidly with complete disappearance of target eDNA within one week (*pers. comm.* Rein Brys & David Halfmaerten). Ponds are further influenced by the activity of domestic and wild animals, which can increase suspended solids within the water column and change the properties of an eDNA sample (Williams, Huyvaert & Piaggio, 2017). These external influences may also transfer eDNA between water bodies and potentially cause false positive detections (Klymus *et al.*, 2017b).

The small and shallow nature of ponds subjects these systems to more extreme conditions than deeper water bodies, including larger fluctuations in temperature range and potentially greater exposure to UV light, although higher turbidity and dense vegetation in some ponds will limit UV light penetration (Kazanjian *et al.*, 2018). Temperature, UV light, and pH all influence eDNA shedding and degradation rates, and can affect the amount of eDNA present within a waterbody (Strickler *et al.*, 2015; Robson *et al.*, 2016; Buxton *et al.*, 2017b; Goldberg *et al.*, 2018). Many ponds are successional in nature and often support abundant emergent and semi-terrestrial vegetation with

substantial (relative to waterbody size) shallow marginal drawdown zones in some cases, creating ideal habitat for multiple invertebrate and amphibian species. As water volume decreases over time, ponds become increasingly ephemeral or seasonal (Wood *et al.*, 2003). Accessing these waters via wet, vegetated margins may make cross-contamination between sites hard to avoid (Biggs *et al.*, 2015), while high levels of organic debris in late succession ponds and duckweed-dominated (*Lemna* spp.) ponds can exacerbate difficulties in collecting clean, debris-free samples.

Crucially, ponds can be highly anoxic due to poor wind-mixing and mass decomposition of terrestrial, submerged, and emergent vegetation, resulting in extremely low oxygen content at the bottom of the water column (Sayer *et al.*, 2013; Kazanjian *et al.*, 2018). Anoxic conditions were shown to slow marine eDNA decay (Weltz *et al.*, 2017), but impacts of anoxia on pond eDNA have not been investigated. Slow decay may affect inferences made from eDNA regarding contemporary species presence; however, anoxic conditions dramatically enhance preservation of pond sediments and the communities that live there, providing information on historical pond biodiversity (Alderton *et al.*, 2017; Emson *et al.*, 2017).

1.7 Thesis rationale

Ponds are a crucial component of freshwater networks, but are poorly represented in catchment-scale legislation, monitoring, and management. Ponds are challenging to monitor due to their high density across landscapes as well as the broad range of taxa they support, both individually and combined. Lack of appropriate monitoring tools has prevented comprehensive, long-term, and systematic assessment of these ecosystems. eDNA analysis could transform pond monitoring through rapid and repeated assessment of individual species or entire communities at the pondscape. While ponds have been included in eDNA research, they have been understudied in comparison to larger lakes or lotic ecosystems. To date, no study has fully investigated the prospects and ecological implications of eDNA monitoring in ponds. This thesis will evaluate the utility of eDNA analysis for monitoring biodiversity associated with ponds in the UK, and address issues pertinent to pond conservation and management. Across all chapters, I will investigate three overarching questions.

1.7.1 Can eDNA analysis be used to monitor threatened biodiversity associated with ponds?

Given the biodiversity value of ponds and the number of rare, protected, and unique species these systems support (Wood *et al.*, 2003; Hill *et al.*, 2018), it is vital to know whether eDNA analysis is an effective monitoring tool in these systems. Targeted eDNA assays have been designed for a number of rare and threatened pond biota, including macroinvertebrates (Thomsen *et al.*, 2012; Doi *et al.*, 2017), amphibians (Thomsen *et al.*, 2012; Goldberg *et al.*, 2018), reptiles (Davy *et al.*, 2015; Kundu *et al.*, 2018; Raemy & Ursenbacher, 2018), fish (Thomsen *et al.*, 2012), and mammals (Thomsen *et al.*, 2012). Similarly, eDNA metabarcoding has been used to assess vulnerable communities in temperate and tropical ponds (Klymus *et al.*, 2017b; Bálint *et al.*, 2018). The threatened *T. cristatus* has been particularly prominent in eDNA research, from conception (Thomsen *et al.*, 2012), validation against conventional tools (Rees *et al.*, 2014a; Biggs *et al.*, 2015), and method development (Buxton *et al.*, 2017a, b, 2018; Rees *et al.*, 2017) to formal recognition and deployment of an eDNA assay for widespread monitoring (Natural England, 2015). *T. cristatus* is also the focus of two chapters of this thesis. In Chapter 2, I compare the sensitivity of eDNA metabarcoding against targeted qPCR for *T. cristatus* detection, and then evaluate eDNA metabarcoding as a tool for ecological hypothesis testing using the *T. cristatus* literature in Chapter 3. In Chapter 4, I validate a novel eDNA assay for another threatened pond species, the crucian carp (*Carassius carassius*), which is one of few fish species associated with ponds. Finally, I investigate whether eDNA metabarcoding of pond water can be used as a tool to monitor distribution of conservation priority mammals (Chapter 6), including otter (*Lutra lutra*), water vole (*Arvicola amphibius*), and beaver (*Castor fiber*).

1.7.2 Can eDNA metabarcoding be used to survey biodiversity at landscape level, including semi-aquatic and terrestrial taxa?

eDNA metabarcoding can generate extensive taxonomic inventories and provide multi-species distribution data to inform management and policy (Lawson Handley, 2015; Deiner *et al.*, 2017). This tool has been used to survey temperate and tropical amphibian communities (Valentini *et al.*, 2016; Bálint *et al.*, 2018) and fish assemblages (Valentini *et al.*, 2016; Evans *et al.*, 2017a; Li *et al.*, 2018b, c) in ponds, and there is potential for

survey of semi-aquatic and terrestrial species (Klymus *et al.*, 2017b; Ushio *et al.*, 2017, 2018) from pond water. I provide evidence that ponds can provide natural samples of biodiversity in the wider environment (Chapters 3 & 6), and demonstrate the power of eDNA metabarcoding to upscale pond biodiversity monitoring and research (Chapter 3). I evaluate eDNA metabarcoding as a tool for invertebrate assessment in ponds (Chapter 5) as well as monitoring semi-aquatic and terrestrial mammals in the wider landscape (Chapter 6).

1.7.3 What are the prospects of eDNA metabarcoding for community investigation in ponds?

Studies have used eDNA metabarcoding to identify biodiversity associated with ponds (Valentini *et al.*, 2016; Evans *et al.*, 2017a; Klymus *et al.*, 2017b; Ushio *et al.*, 2017, 2018; Bálint *et al.*, 2018; Li *et al.*, 2018b, c), but an ecological perspective was not applied to the species inventories generated. Metabarcoding could enable analyses of ecological communities on a deeper level, including multi-species occupancy, diversity metrics, species interactions, ecological networks (e.g. habitats, trophic, pollination), and biomonitoring (Deiner *et al.*, 2017). In this thesis, I apply several of the aforementioned analyses to eDNA metabarcoding data and somewhat lessen this knowledge gap in the eDNA literature. I evaluate eDNA metabarcoding as a tool for ecological hypothesis testing, specifically biotic (species associations) and abiotic determinants of *T. cristatus* occupancy and vertebrate species richness (Chapter 3). I use eDNA metabarcoding to assess alpha and beta diversity of invertebrate communities in relation to fish stocking of ponds (Chapter 5). Finally, I examine spatiotemporal variation in the vertebrate communities of ponds, with focus on semi-aquatic and terrestrial mammals that visit these water bodies (Chapter 6).

1.8 Data chapter summaries

This section summarises the aims of each chapter and their contribution to the overarching questions identified above.

1.8.1 Chapter 2: Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*)

In Chapter 2 (published in *Ecology and Evolution*), I perform a large-scale comparison (*N* = 532 ponds) of qPCR and metabarcoding sensitivity for detection of the threatened *T. cristatus*. eDNA samples were previously analysed for *T. cristatus* by commercial companies using a qPCR assay designed by Thomsen *et al*. (2012). The samples were screened again by eDNA metabarcoding using vertebrate-specific primers to obtain community composition alongside *T. cristatus* presence-absence. Detection and signal strength of *T. cristatus* eDNA by metabarcoding are hypothesised to be comparable to qPCR. eDNA metabarcoding is expected to provide distribution data on wider vertebrate biodiversity present without compromising single-species detection.

1.8.2 Chapter 3: Ground-truthing environmental DNA (eDNA) metabarcoding for ecological hypothesis testing at the pondscape

In Chapter 3 (available as a pre-print on *bioRxiv* and submitted to *Environmental DNA*), I use the community data generated by eDNA metabarcoding in Chapter 2 to assess the utility of this tool for ecological hypothesis testing at the pondscape, with *T. cristatus* as a focal species. Specifically, I compare determinants of eDNA-based *T. cristatus* occurrence to those reported in the existing literature, and identify determinants of vertebrate species richness. I also evaluate the appropriateness of the *T. cristatus* Habitat Suitability Index (HSI) to predict eDNA-based *T. cristatus* occupancy and vertebrate species richness. I hypothesise that *T. cristatus* occupancy will be lower in ponds containing fish (particularly the predatory three-spined stickleback [*Gasterosteus aculeatus*]), and waterfowl, but higher in ponds with other amphibians, namely the smooth newt (*Lissotriton vulgaris*). Pond density, water quality, woodland, grassland, and HSI score are expected to positively influence *T. cristatus* occupancy, whereas pond area, macrophyte cover, and canopy cover are expected to have negative effects. Vertebrate species richness is hypothesised to increase with pond density, terrestrial habitat quality, and HSI score, but decrease with macrophyte and canopy cover.

1.8.3 Chapter 4: Development and application of environmental DNA surveillance for the threatened crucian carp (*Carassius carassius*)

In Chapter 4 (published in *Freshwater Biology*), I design and validate a qPCR assay to target eDNA from the threatened *C. carassius*, one of few fish associated with ponds in the UK. I sampled water from 10 ponds with *C. carassius* at different densities (confirmed by fyke netting), and 10 ponds without fish. I compare species detection by eDNA analysis to fyke netting, and evaluate whether the qPCR assay can estimate relative abundance of *C. carassius*. I also identify biotic and abiotic factors that influence eDNA detection using a hierarchical occupancy model, and which of these also affect eDNA quantification using a mixed effects model. I hypothesise that: eDNA analysis will be comparable to fyke netting for *C. carassius* presence-absence; eDNA concentration will increase as a function of conventional density estimation; and *C. carassius* density, temperature, pH, conductivity, surface dissolved oxygen, macrophyte cover and tree shading will affect eDNA detection and quantification.

1.8.4 Chapter 5: Assessing the impact of the threatened crucian carp (*Carassius carassius*) on pond invertebrate diversity - a comparison of conventional and molecular tools

In Chapter 5 (in preparation for submission to journal), I compare invertebrate diversity in ponds stocked with *C. carassius* for conservation purposes and fishless ponds using standard sweep-netting and microscopy alongside metabarcoding (DNA and eDNA). Invertebrate samples were collected from the same ponds sampled for eDNA in Chapter 4. The invertebrate samples were processed and tissue DNA extracted for DNA metabarcoding. Data produced by each method are examined individually and in combination to assess the impact of *C. carassius* on invertebrate diversity. I hypothesise ponds with *C. carassius* will have reduced alpha diversity, but beta diversity of ponds will be enhanced due to greater community heterogeneity induced by the different taxa present in ponds with or without *C. carassius*. Sweep-netting with microscopy and DNA metabarcoding are expected to generate highly similar taxonomic inventories, whereas eDNA metabarcoding is expected to detect taxa not found by the other approaches. Both DNA and eDNA metabarcoding are anticipated to provide species-level identification for

specimens that can only be identified to genus- or family-level by sweep-netting and microscropy.

1.8.5 Chapter 6: Environmental DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority mammals

In Chapter 6 (available as a pre-print on *bioRxiv* and submitted to *Biological Conservation*), I evaluate the potential of ponds to provide natural samples of biodiversity in the wider environment. Specifically, I examine the capacity of eDNA in ponds to reveal semi-aquatic and terrestrial mammals present in the surrounding area. I focus on nine mammal species of conservation or management concern in the UK, which require non-invasive monitoring tools to improve species distribution maps. I sampled water bodies in animal enclosures at two wildlife parks in conjunction with behavioural observation of captive animals, followed by sampling of natural ponds at locations across the UK in conjunction with camera trapping and field signs. eDNA metabarcoding is hypothesised to perform better for semi-aquatic mammals than terrestrial species, as eDNA from semi-aquatic species will be evenly distributed in the water column as opposed to localised distribution of eDNA from terrestrial species. The eDNA signal produced by metabarcoding (i.e. read counts) is expected to be higher for species that exhibit behaviours directly involving water. eDNA metabarcoding is hypothesised to detect more mammal species than camera trapping and field signs.

Chapter 2: Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*)

Larval great crested newt (*Triturus cristatus*) (Laurenti, 1768)
© user: Sam Dredge | Flickr | CC BY-NC-ND 2.0

This chapter has been published as

Harper, L.R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H.C., Gough, K.C., Lewis, E., Adams, I.P., Brotherton, P., Phillips, S. & Hänfling, B (2018) Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*). *Ecology and Evolution*, **8**, 6330–6341.

Abstract

Environmental DNA (eDNA) analysis is a rapid, cost-effective, non-invasive biodiversity monitoring tool which utilises DNA left behind in the environment by organisms for species detection. The method is used as a species-specific survey tool for rare or invasive species across a broad range of ecosystems. Recently, eDNA and 'metabarcoding' have been combined to describe whole communities rather than focusing on single target species. However, whether metabarcoding is as sensitive as targeted approaches for rare species detection remains to be evaluated. The great crested newt (*Triturus cristatus*) is a flagship pond species of international conservation concern and the first UK species to be routinely monitored using eDNA. We evaluate whether eDNA metabarcoding has comparable sensitivity to targeted real-time quantitative PCR (qPCR) for *T. cristatus* detection. Extracted eDNA samples (N = 532) were screened for *T. cristatus* by qPCR and analysed for all vertebrate species using high-throughput sequencing technology. With qPCR and a detection threshold of 1 of 12 positive qPCR replicates, newts were detected in 50% of ponds. Detection decreased to 32% when the threshold was increased to 4 of 12 positive qPCR replicates. With metabarcoding, newts were detected in 34% of ponds without a detection threshold, and in 28% of ponds when a threshold (0.028%) was applied. Therefore, qPCR provided greater detection than metabarcoding, but metabarcoding detection with no threshold was equivalent to qPCR with a stringent detection threshold. The proportion of *T. cristatus* sequences in each sample was positively associated with the number of positive qPCR replicates (qPCR score) suggesting eDNA metabarcoding may be indicative of eDNA concentration. eDNA metabarcoding holds enormous potential for holistic biodiversity assessment and routine freshwater monitoring. We advocate this community approach to freshwater monitoring to guide management and conservation, whereby entire communities can be initially surveyed to best inform use of funding and time for species-specific surveys.

2.1 Introduction

Species monitoring has rapidly evolved with the advent of environmental DNA (eDNA) analysis (Lawson Handley, 2015). eDNA analysis allows highly sensitive detection of rare and invasive species and is increasingly used for surveys of aquatic species (Thomsen *et al.*, 2012; Biggs *et al.*, 2015; Davy *et al.*, 2015; Smart *et al.*, 2016; Evans *et al.*, 2017b). This non-invasive approach uses intra- and extracellular DNA (e.g. mucus, skin cells, hair, urine, faeces, gametes, deceased remains) released into the environment by organisms to survey for species and assess their distribution (Rees *et al.*, 2014b; Lawson Handley, 2015; Goldberg *et al.*, 2016). Typically for eDNA analysis, DNA is extracted from environmental samples (water, soil, air) and analysed using a targeted or passive approach. The targeted approach uses species-specific primers with conventional PCR (PCR), real-time quantitative PCR (qPCR), or droplet digital PCR (ddPCR), to determine presence-absence and estimate abundance of single species (Shaw, Weyrich & Cooper, 2016b; Goldberg *et al.*, 2016). Conversely, the passive approach uses conserved primers (i.e. primers with binding sites that are shared across multiple taxa, and flank a region of highly variable DNA sequence that enables discrimination between these taxa) and PCR to sequence whole communities with high-throughput sequencing, termed eDNA metabarcoding (Taberlet *et al.*, 2012; Shaw *et al.*, 2016b; Valentini *et al.*, 2016; Deiner *et al.*, 2017). Passive eDNA monitoring is particularly attractive to ecologists for biodiversity assessment as a means to detect entire species assemblages alongside rare or invasive species (Lacoursière-Roussel *et al.*, 2016a; Blackman *et al.*, 2017). However, this gain in community understanding may come at the cost of accuracy and sensitivity. Direct comparisons of these two approaches are essential to determine whether they have comparable power and yield similar results.

Although in its relative infancy, eDNA metabarcoding has proven effective for community biodiversity assessment across a range of taxa in varying environments, particularly freshwater herpetofauna and fish (Civade *et al.*, 2016; Hänfling *et al.*, 2016; Lacoursière-Roussel *et al.*, 2016a; Lopes *et al.*, 2016; Shaw *et al.*, 2016a; Valentini *et al.*, 2016; Evans *et al.*, 2017a; Bálint *et al.*, 2018). However, eDNA metabarcoding is confounded by potential amplification bias during PCR, preventing capture of all species present in a given area (Kelly *et al.*, 2014). Species' DNA in community samples is also in competition to bind to metabarcoding primers during PCR, where more common templates are more likely to be amplified. High abundance species may thus prevent

detection of low abundance species, whether by fewer individuals or less DNA shed, resulting in 'species masking' (Kelly *et al.*, 2014; Brandon-Mong *et al.*, 2015; Evans *et al.*, 2016b). eDNA metabarcoding may therefore be less capable of identifying eDNA from rare species within a community than species-specific qPCR (Evans *et al.*, 2016b).

The sensitivity of eDNA metabarcoding has been evaluated against conventional biodiversity monitoring methods in freshwater ecosystems (Civade *et al.*, 2016; Hänfling *et al.*, 2016; Lopes *et al.*, 2016; Shaw *et al.*, 2016a; Valentini *et al.*, 2016; Evans *et al.*, 2017a; Bálint *et al.*, 2018), yet specific investigations comparing the sensitivity of eDNA metabarcoding and targeted qPCR are sparse. Similarly, comparisons of qPCR and conventional survey for species monitoring have included cost projections (Biggs *et al.*, 2014; Davy *et al.*, 2015; Smart *et al.*, 2016; Evans *et al.*, 2017b), but cost has not been thoroughly assessed in qPCR and eDNA metabarcoding comparisons (Lacoursière-Roussel *et al.*, 2016a; Schneider *et al.*, 2016). Schnieder *et al.* (2016) achieved improved detection of invasive mosquito species with qPCR and eDNA metabarcoding as opposed to conventional sampling. Although qPCR provided higher detection probability for two species, metabarcoding achieved comparable results for a third species and allowed simultaneous detection of invasive mosquito species and other taxa in a single sequencing run without development of multiple species-specific markers. In another study, eDNA metabarcoding failed to detect wood turtle (*Glyptemys insculpta*) in four rivers where qPCR and conventional visual survey detected the species (Lacoursière-Roussel *et al.*, 2016a). Amplification of longer fragments during metabarcoding versus qPCR could account for difference in sensitivity of the two methods, with the shorter qPCR assay being more capable of detecting heavily degraded DNA (Lacoursière-Roussel *et al.*, 2016a). Further research is clearly needed to determine whether these two approaches are comparable.

The great crested newt (*Triturus cristatus*, Fig. 2.1) is a model organism for eDNA-based monitoring. *T. cristatus* secrete mucus, breed in water, and produce aquatic eggs and larvae – all sources of DNA deposition in ponds. The species is rare in parts of the UK and Europe, and as such, all life stages are protected by UK and European legislation (Rees *et al.*, 2014a; Buxton *et al.*, 2017b). eDNA analysis using targeted qPCR has been repeatedly verified against conventional surveying (bottle trapping, torchlight counts, larval netting, egg searches) for *T. cristatus* and found to

Figure 2.1: Adult male great crested newt (*Triturus cristatus*). Photograph by Brett Lewis (Lewis Ecology, Brett Lewis Photography).

achieve comparable or improved species detection (Thomsen *et al.*, 2012; Rees *et al.*, 2014a; Biggs *et al.*, 2015). eDNA sampling can be undertaken with relative ease, is cost-efficient (Biggs *et al.*, 2014), and can be implemented in large-scale citizen science monitoring programmes without loss of species detection (Biggs *et al.*, 2015). *T. cristatus* is the first species to be routinely monitored using eDNA in the UK (Natural England, 2015) and targeted eDNA assays are now offered as a commercial service by ecological consultancies. The targeted eDNA assay is highly effective for *T. cristatus* detection; however, should metabarcoding have comparable sensitivity, this approach would allow detection of *T. cristatus* alongside pond communities and potentially enable more cost-effective monitoring of entire ecosystems and ecological hypothesis testing.

Here, we perform a large-scale comparison (N = 532 ponds) of eDNA metabarcoding and targeted qPCR for *T. cristatus* detection to compare method sensitivity. A single primer pair that is vertebrate-specific for mitochondrial DNA (mtDNA) and requires no *a priori* knowledge of species composition, was employed for eDNA metabarcoding. The metabarcoding results were then compared to results obtained using the standard *T. cristatus* qPCR assay (Biggs *et al.*, 2015). Our hypotheses are as

follows: (1) eDNA metabarcoding will give equivalent results to qPCR for *T. cristatus* detection, (2) eDNA metabarcoding sequence read count for *T. cristatus* will increase as qPCR score (the number of positive qPCR replicates) increases, indicative of eDNA concentration, and (3) metabarcoding primers will amplify DNA from all taxa equally well and no bias towards amplification of *T. cristatus* will occur (bias would be indicated by a positive association between the proportion of *T. cristatus* sequence reads and PCR product concentration). We also examined cost and investigator effort required by each approach to determine whether a trade-off between cost, time, and amount of data generated exists.

2.2 Materials and methods

2.2.1 Sampling

Samples from 532 ponds distributed across three UK counties (Cheshire, Kent and Lincolnshire) were analysed for this project. Of these, 508 ponds (ranging from 9 to 9375 m^2) were sampled as part of *T. cristatus* surveys through Natural England's Great Crested Newt Evidence Enhancement Programme. *T. cristatus* egg searches were performed once during the daytime at 506 of 508 ponds. Any other life stages seen were also recorded. A further 24 ponds were sampled for eDNA by ecological consultants for private contracts but egg searches were not undertaken. All water samples were collected using methodology outlined by Biggs *et al*. (2015). Water samples were then sent to Fera Science Ltd (Natural England) and ADAS (private contracts), where one eDNA sample per pond was produced and analysed according to laboratory protocols established by Biggs *et al.* (2015). Details of sampling methodology and laboratory protocols are provided in Appendix 2.1.

2.2.2 Targeted qPCR for *T. cristatus*

Targeted qPCR was conducted as part of the *T. cristatus* monitoring programmes mentioned above in Fera Science Ltd and ADAS laboratories during 2015. Both laboratories used a standardised protocol, which tests for PCR inhibitors and sample degradation prior to testing for *T. cristatus* (Biggs *et al.*, 2015). Extracted DNA was

amplified by TaqMan probe qPCR using published primers and probe (Thomsen *et al.*, 2012) to amplify an 81 bp fragment of the cytochrome *b* gene. For each sample, 12 qPCR replicates were performed and a sample recorded as positive for *T. cristatus* if one or more qPCR replicates were positive. Following qPCR, the eDNA samples were placed in storage at -80 °C.

2.2.3 Metabarcoding of vertebrate communities

eDNA samples were stored at -20 °C until PCR amplification. Metabarcoding was performed using published vertebrate-specific primers (Riaz *et al.*, 2011) which amplify a 73-110 bp fragment of the 12S ribosomal RNA (rRNA) gene. The assay was first validated *in silico* using ecoPCR software (Ficetola *et al.*, 2010) against a custom, phylogenetically curated reference database for UK vertebrates. Full details of reference database construction are provided in Appendix 2.1. The complete reference database compiled in GenBank format has been deposited in a dedicated GitHub repository for this chapter, permanently archived at: https://doi.org/10.5281/zenodo.2633978. Parameters set allowed a 50-250 bp fragment and maximum of three mismatches between the primer pair and each sequence in the reference database. Primers were then validated against tissue DNA extracted from UK amphibian species (Appendix 2.1) having been previously validated *in vitro* for UK fish communities by Hänfling *et al.* (2016). After primer validation, a two-step PCR protocol was used to construct metabarcoding libraries from the eDNA samples. During the first PCR, the target region was amplified using metabarcoding primers, comprised of the aforementioned specific locus primer, random hexamers, sequencing primer, and pre-adapter (Illumina, 2011). DNA (0.284 ng/µL) from the cichlid *Rhamphochromis esox* was used for PCR positive controls (six per PCR plate; *n* = 114), whilst sterile molecular grade water (Fisher Scientific UK Ltd, UK) substituted template DNA for No Template Controls (NTCs, six per PCR plate; *n* = 114). In the second PCR, Multiplex Identification (MID) tags (unique 8-nucleotide sequences) and Illumina MiSeq adapter sequences were added to the amplified product. Two independent libraries were constructed, each containing 266 eDNA samples, 57 NTCs, and 57 positive controls. Sequencing was performed on an Illumina MiSeq using 2 x 300 bp V3 chemistry (Illumina, Inc, CA, USA) at Fera Science Ltd. The first sequencing run revealed human (*Homo sapiens*) contamination across samples and in some PCR controls; therefore, reactions prepared for the second sequencing run were sealed with mineral oil to minimise

PCR contamination. Full details of the eDNA metabarcoding workflow are provided in Appendix 2.1.

2.2.4 Bioinformatic processing

Illumina data was converted from raw sequences to taxonomic assignment using a custom pipeline for reproducible analysis of metabarcoding data: metaBEAT (metaBarcoding and eDNA Analysis Tool) v0.8 (https://github.com/HullUni-bioinformatics/metaBEAT). Bioinformatic data processing/analysis largely followed the workflow outlined by Hänfling *et al.* (2016), with minor modifications (see Appendix 2.1 for details). To ensure reproducibility of analyses, the workflow has been deposited in the GitHub repository.

2.2.5 Data analysis

All downstream analyses were performed in the statistical programming environment R v.3.3.2. (R Core Team, 2017). Data and R scripts have been deposited in the GitHub repository. Manipulation of the dataset produced by metaBEAT is described in Appendix 2.1.

2.2.5.1 Detection thresholds and contamination

At present, there are no standard guidelines for eDNA analysis to indicate the minimum number of positive eDNA samples or replicates required to class sites as species positive (Goldberg *et al.*, 2016). Samples analysed by qPCR in this study were previously considered *T. cristatus* positive if one or more qPCR replicates gave a positive result (Biggs *et al.*, 2015). We term this analysis qPCR NT (No Threshold). This inference of species presence is employed across many studies but may not be reliable or reproducible (Goldberg *et al.*, 2016). More stringent qPCR thresholds reduced detection sensitivity for palmate newt (*Lissotriton vulgaris*) (Smart *et al.*, 2016), but may be necessary to ensure consistency and prevent false positives (Rees *et al.*, 2014b). To facilitate comparison with current qPCR scoring (our NT interpretation) and eDNA metabarcoding, we applied a stringent qPCR threshold of $\geq$ 4/12 positive qPCR replicates to infer species presence, and termed the new analysis qPCR TA (Threshold Applied).

The raw eDNA metabarcoding dataset with no detection thresholds applied was

termed metabarcoding NT (No Threshold). A second dataset was constructed to reduce the potential for false positives by application of a species-specific threshold: a species was only classed as present at a given site if its sequence frequency exceeded a species-specific threshold. Thresholds for each species were defined by analysing sequence data from PCR positive controls (n = 114) and identifying the maximum sequence frequency for a given species across all PCR positive controls (Table S2.2). For example, the species-specific false positive sequence threshold for *T. cristatus* was 0.028% to omit all false detections in the PCR positive controls. The resultant dataset was termed metabarcoding TA (Threshold Applied).

We tested whether mineral oil reduced contamination by analysing the distribution of positive control sequences (*R. esox*) and *H. sapiens* DNA in eDNA samples, and any DNA in NTCs, across both sequencing runs using binomial Generalized Linear Mixed Models (GLMMs) within the R package lme4 v1.1-12 (Bates *et al.*, 2015). The response variable was presence-absence of contamination and explanatory variables were PCR plate (random effect) and sequencing run, i.e. mineral oil sealed versus non-sealed (fixed effect). *H. sapiens* DNA may be present in eDNA samples as a real environmental signal or contaminant prior to PCR and thus may not be a true PCR contaminant. Consequently, contamination in eDNA samples was examined using several model permutations, where contamination comprised both *R. esox* and *H. sapiens* DNA, *R. esox* DNA alone, and *H. sapiens* DNA alone. An information-theoretic approach using Akaike's Information Criteria (AIC) to evaluate model fit was employed, where low AIC models are more parsimonious than high AIC models (Akaike, 1973). Significance of the fixed effect in the model was tested by a Likelihood Ratio Test (LRT).

2.2.5.2 Comparison of eDNA methods for T. cristatus detection

We tested the null hypothesis of no significant difference in sensitivity of qPCR and metabarcoding. Overall agreement between eDNA metabarcoding and qPCR for *T. cristatus* detection was measured using Cohen's kappa coefficient (Cohen, 1960), following which Pearson's Chi-squared Test for Independence was used to test equality of *T. cristatus* detection between eDNA approaches.

Previously, Biggs *et al.* (2015) found qPCR score was an inconsistent predictor of *T. cristatus* abundance, where ponds with low scores had low newt counts but high scores did not correspond to large populations. qPCR score may only be proxy for the

amount of DNA present rather than the number of individuals. The relationship between read count and qPCR score has not been examined previously, and whether read production is indicative of DNA concentration remains unknown. We hypothesised samples with higher qPCR score would have increased *T. cristatus* read count. First, the average number of *T. cristatus* reads produced by eDNA metabarcoding per qPCR score (1-12 of 12) was calculated. A Spearman Rank Correlation was then used to test for a relationship between average read count and qPCR score.

Following data exploration (see Appendix 2.1), a negative binomial GLMM was used to counter overdispersion and improve model fit. The GLMM examined read count in relation to qPCR score, accounting for other variables that may affect metabarcoding signal strength. Variation in *T. cristatus* read count was examined using the proportion of *T. cristatus* reads within the total number of reads produced for each eDNA sample as the response variable. Sequencing run and PCR plate were considered random effects and all other explanatory variables as fixed effects (qPCR score, sample degradation, sample inhibition, post-PCR eDNA concentration). Presence-absence of sample degradation and inhibition was determined by qPCR in 2015 using methodology outlined by Biggs *et al.* (2015). Model fit was again evaluated using AIC and significance of fixed effects in the model was tested with stepwise backward deletion of terms from the model informed by LRTs. All values were bound in a new data frame and model results plotted for evaluation using the R package ggplot2 v 2.1.0 (Wickham, 2016).

2.2.5.3 Cost and investigator effort

Cost of materials and investigator effort and salary (hourly rate of £21.20 assumed) were calculated for eDNA samples; however, estimates do not include travel to sampling sites, procedural controls, qPCR standards, or consumables and reagents required for assay optimisation. Time required to perform PCR for metabarcoding and qPCR was estimated assuming available machinery to run four PCR plates in parallel and one qPCR plate.

2.3 Results

2.3.1 Targeted qPCR and egg searches

Targeted qPCR detected *T. cristatus* in 253 (49.80%) samples analysed by Fera Science Ltd (*n* = 508). Of 255 (50.20%) samples that were negative, one was inhibited and nine were degraded. qPCR and egg searches produced consistent results for 297 (58.47%) ponds, with 51 (10.04%) positive and 246 (48.43%) negative ponds by both methods. Of the 211 ponds where there was disagreement between methods, 202 (39.76%) were qPCR positive but negative by egg searches, and 7 (1.38%) were positive with egg searches but qPCR negative. Of 24 samples analysed by ADAS, 12 (50.00%) were qPCR negative and 12 (50.00%) were qPCR positive for *T. cristatus*. No egg search data were available for these ponds.

2.3.2 Vertebrate metabarcoding

The *in silico* and *in vitro* primer validation confirmed that *T. cristatus*, and other native UK amphibians tested, can be reliably amplified and identified with the chosen assay (Appendix 2.2, Fig. S2.1). Furthermore, the *in silico* approach showed that the majority of all UK vertebrates can be amplified (see Appendix 2.2 for details). Both sequencing runs had comparable yield and sequencing quality score; summary statistics for each sequencing run and read counts for taxonomic assignment levels are provided in Appendix 2.2 (Tables S2.3, S2.4). A full summary of sequence read count data is given in the archived GitHub repository for this chapter (https://doi.org/10.5281/zenodo.2633978). eDNA metabarcoding identified a combined total of 60 species (Appendix 2.2, Fig. S2.2) across both sequencing libraries, with 375,954 and 508,879 sequences assigned to *T. cristatus* from each library. Analyses of overall pond species compositions inferred by eDNA metabarcoding (Appendix 2.2, Fig. S2.3, Table S2.5) are reported separately (see Chapter 3 and Harper *et al.*, 2018b).

All samples (*N* = 532) were sequenced and of 57 samples that did not produce visible PCR bands, nine generated sequence reads. Notably, the 57 samples were not inhibited or degraded at time of qPCR. Weak PCR bands were observed in some NTCs; therefore, all PCR controls were sequenced (Appendix 2.2, Figs. S2.4-2.6). Six NTCs contained *T. cristatus* DNA but only one exceeded 100 *T. cristatus* reads (307/330 reads).

Twelve other sources occurred in NTCs (Appendix 2.2, Table S2.6); seven occurred in more than one NTC and eight had high maximum read counts (> 100 reads).

Contamination of NTCs (any DNA) and environmental samples (*R. esox*/*H. sapiens* DNA) was observed (Appendix 2.2, Figs. S2.4-2.6). Read counts of NTC contaminants were reduced between sequencing runs with the addition of mineral oil to PCR reactions included on the second sequencing run (Appendix 2.2, Figs. S2.4-2.6) but this reduction was not statistically significant (GLMM: $\chi^2_1 = 2.083$, $F_1 = 1.941$, $P = 0.149$). Mineral oil did not reduce *H. sapiens* DNA signal in environmental samples between sequencing runs either (GLMM: $\chi^2_1 = 3.608$, $F_1 = 3.591$, $P = 0.058$); however, it did reduce *H. sapiens* DNA in combination with cichlid DNA (GLMM: $\chi^2_1 = 10.348$, $F_1 = 21.143$, $P = 0.001$), and cichlid DNA contamination alone (GLMM: $\chi^2_1 = 5.053$, $F_1 = 6.978$, $P = 0.025$) of environmental samples.

2.3.3 eDNA metabarcoding vs. qPCR for *T. cristatus* detection

T. cristatus detection by metabarcoding NT (34.21%) was less sensitive than qPCR NT (49.81%) but marginally higher than qPCR TA (32.71%) ($N = 532$ ponds, Fig. 2.2). Metabarcoding TA had lower detection efficiency (28.01%) and failed to detect *T. cristatus* in 116 and 25 ponds where the species was detected by qPCR NT and qPCR TA respectively. Nonetheless, both molecular approaches attained higher *T. cristatus* detection than daytime egg searches (11.46%) in 506 ponds where all three approaches were implemented.

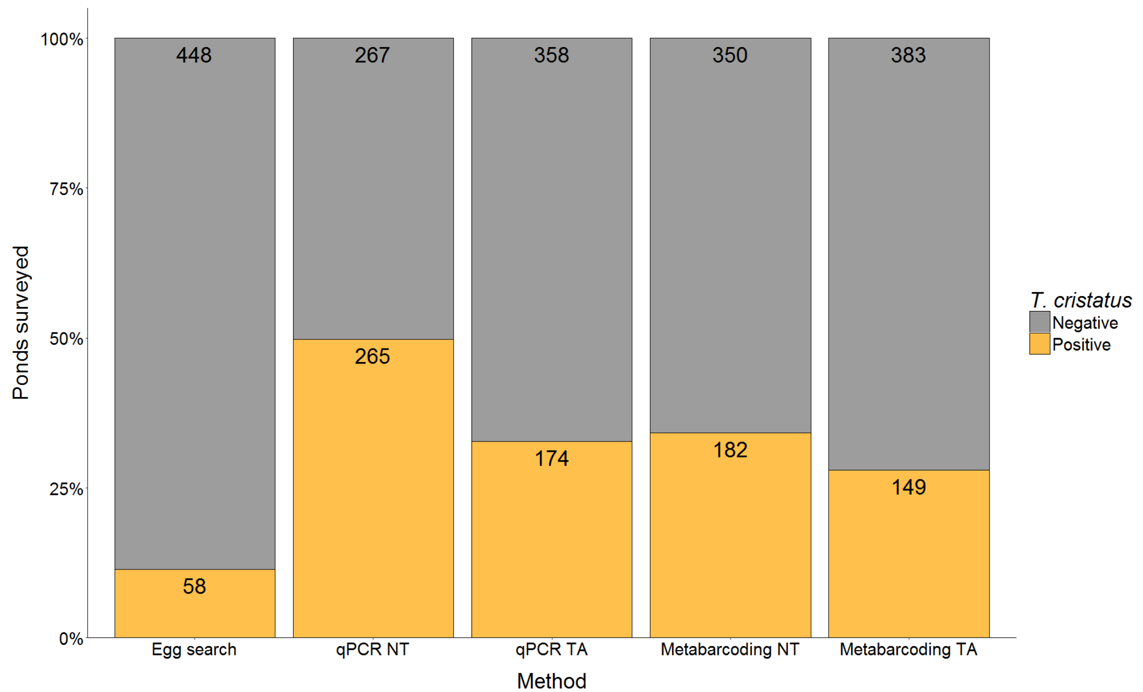


Figure 2.2: Comparison of survey methodology for *T. cristatus* detection in freshwater ponds across the UK. Bars represent proportion of positive and negative *T. cristatus* ponds by each method with frequency displayed on bars.

Overlap between survey methods for positive *T. cristatus* ponds (n = 277), and unique detections by each method are summarised in Fig. 2.3. Negative *T. cristatus* ponds (n = 229) are examined in combination with species positive ponds in Appendix 2.2 (Table S2.7). Each survey method detected the species in ponds where other methods failed. Despite lower *T. cristatus* detection efficiency, egg searches detected the species in six ponds where it went undetected by qPCR and metabarcoding. Metabarcoding NT and metabarcoding TA revealed *T. cristatus* in seven ponds which other methods did not, whilst qPCR NT and qPCR TA detected *T. cristatus* in 33 ponds unique to other methods. All methods detected *T. cristatus* in 32 ponds, and both metabarcoding and qPCR identified *T. cristatus* in 86 ponds. Disagreement between molecular methods was more likely when samples were positive rather than negative by qPCR. Without thresholds, 39.25% of qPCR positive ponds (n = 265) were negative by metabarcoding, but 7.87% of qPCR negative ponds (n = 267) were positive by metabarcoding. With thresholds, 29.31% of qPCR positive ponds (n = 174) were negative by metabarcoding, whereas 7.26% of qPCR negative ponds (n = 358) were positive by metabarcoding.

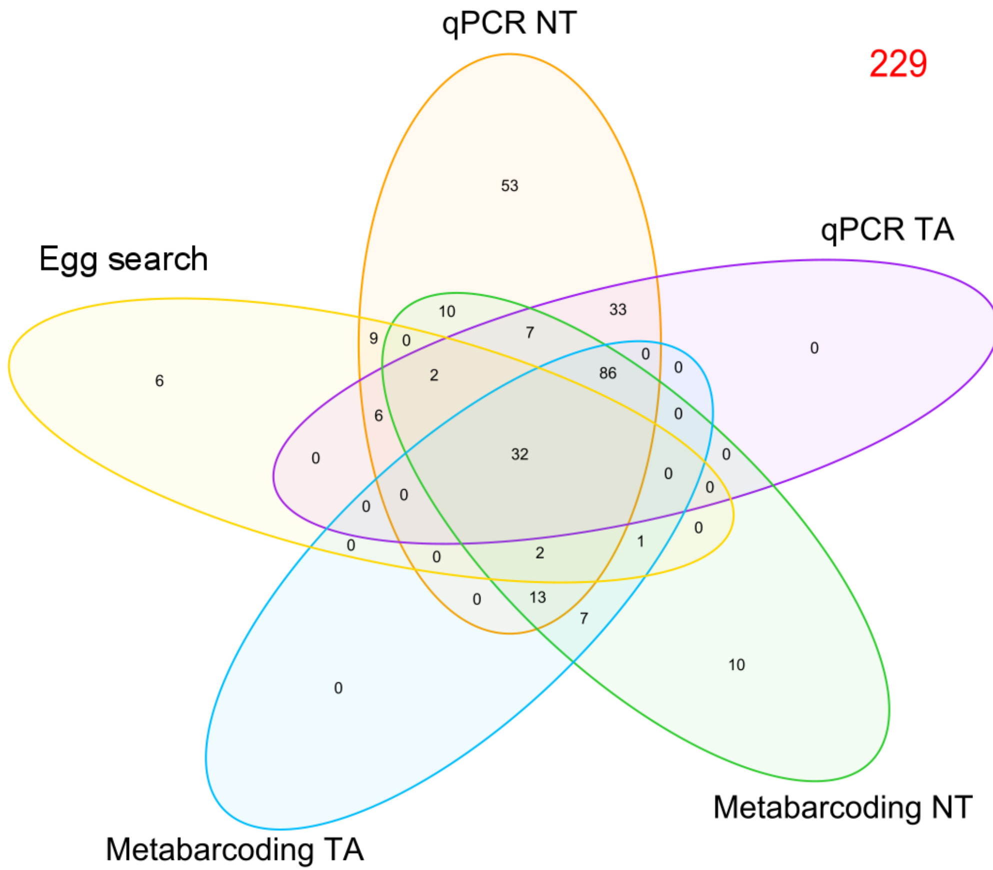


Figure 2.3: Venn diagram which summarises the number of positive *T. cristatus* detections (n = 277) by each method (egg search, qPCR NT, qPCR TA, metabarcoding NT, and metabarcoding TA), and overlap in *T. cristatus* detection between methods for 506 ponds where all methods were applied. Negative *T. cristatus* detections (n = 229) are highlighted in red.

Agreement between eDNA approaches is summarised in Table 2.1. Agreement was strongest between eDNA approaches when the qPCR detection threshold was applied, irrespective of whether the metabarcoding detection threshold was applied. Metabarcoding (NT or TA) and qPCR TA did not significantly differ in their detection of *T. cristatus* (Table 2.1).

Table 2.1: Summary of analyses testing for agreement between eDNA approaches, with and without thresholds, for *T. cristatus* detection. Cohen's kappa coefficient (k) represents strength of agreement between methods (1 = 100%). Pearson's Chi-squared Test for Independence tested whether methods significantly differed for *T. cristatus* detection.

Comparison	Probability of observed agreement	Probability of expected agreement	k	Overall agreement	χ^2	DF	*P*
Metabarcoding NT qPCR NT	0.77	0.50	0.53	Moderate	25.94	1	3.521×10^{-7}
Metabarcoding TA qPCR NT	0.74	0.50	0.48	Moderate	52.291	1	4.787×10^{-13}
Metabarcoding NT qPCR TA	0.84	0.56	0.63	Good	0.207	1	0.649
Metabarcoding TA qPCR TA	0.86	0.58	0.66	Good	2.561	1	0.110

An identical positive correlation was observed between qPCR score and the average number of *T. cristatus* reads obtained for samples belonging to each qPCR score (r_s = 0.648, df = 11, $P = 0.020$), regardless of threshold application to the metabarcoding data. Despite some inconsistency across qPCR scores, samples with a higher qPCR score generally had more *T. cristatus* reads, supportive of a relationship between metabarcoding and abundance of eDNA from single species. Notably, metabarcoding produced *T. cristatus* reads for qPCR NT and qPCR TA negative samples, but the *T. cristatus* metabarcoding signal of these (qPCR $NT_{negative}$ = 2639 reads max., qPCR $TA_{negative}$ = 3075 reads max.) was much lower than samples with higher qPCR score (max. 65,325 reads; Appendix 2.5). Further examination of the relationship between qPCR score and metabarcoding TA revealed qPCR score and post-PCR eDNA concentration of samples also influenced the proportion of *T. cristatus* reads, i.e. relative *T. cristatus* sequence read production (Table 2.2).

Table 2.2: Summary of analyses testing for variation in proportion of *T. cristatus* sequence reads in a sample produced by eDNA metabarcoding, attributable to qPCR score or post-PCR eDNA concentration. Test statistic is for LRT used.

Model variables	*N* (ponds)	DF	AIC	Effect size	Standard error	χ^2	*F*	*P*
qPCR score	532	1	1578.3	0.373	0.032	150.682	147.117	< *0.001*
post-PCR eDNA concentration	532	1	1441.9	-0.056	0.015	14.272	12.457	< *0.001*

A significant positive relationship was observed between qPCR score and the proportion of *T. cristatus* reads within total reads per sample ($P < 0.001$) (Fig. 2.4a). Conversely, post-PCR eDNA concentration had a significant negative influence on the proportion of *T. cristatus* reads ($P < 0.001$), where read proportion decreased as post-PCR eDNA concentration increased (Fig. 2.4b).

2.3.4 Comparison of method cost and investigator effort

Cost and investigator effort for both eDNA approaches were comparable. Metabarcoding was marginally more expensive (£3 per pond) than qPCR, but used 1 day less of investigator effort. A full breakdown of expenditure per pond is given in the archived GitHub repository for this chapter (https://doi.org/10.5281/zenodo.2633978) and summarised in Fig. 2.5.

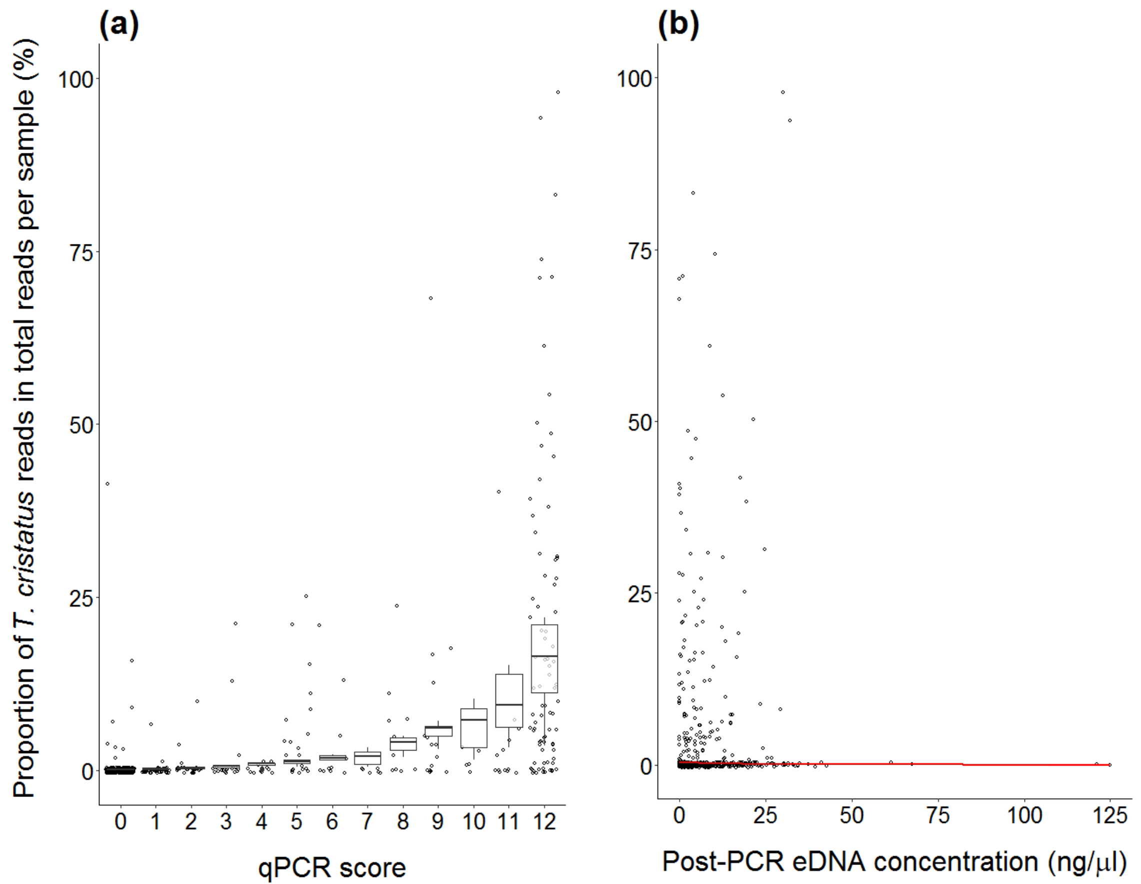


Figure 2.4: Relationship between fixed effects (qPCR score, post-PCR eDNA concentration) and response variable (proportion of *T. cristatus* reads) in eDNA samples, as predicted by the negative binomial GLMM. The 95% CIs, as calculated using the predicted proportions, and standard error for these predictions are given for each relationship. The observed data (points) are also displayed against the predicted relationships (boxes, line). The proportion of *T. cristatus* reads within eDNA samples increased as qPCR score increased **(a)**. Conversely, the proportion of *T. cristatus* reads decreased as post-PCR eDNA concentration increased **(b)**.

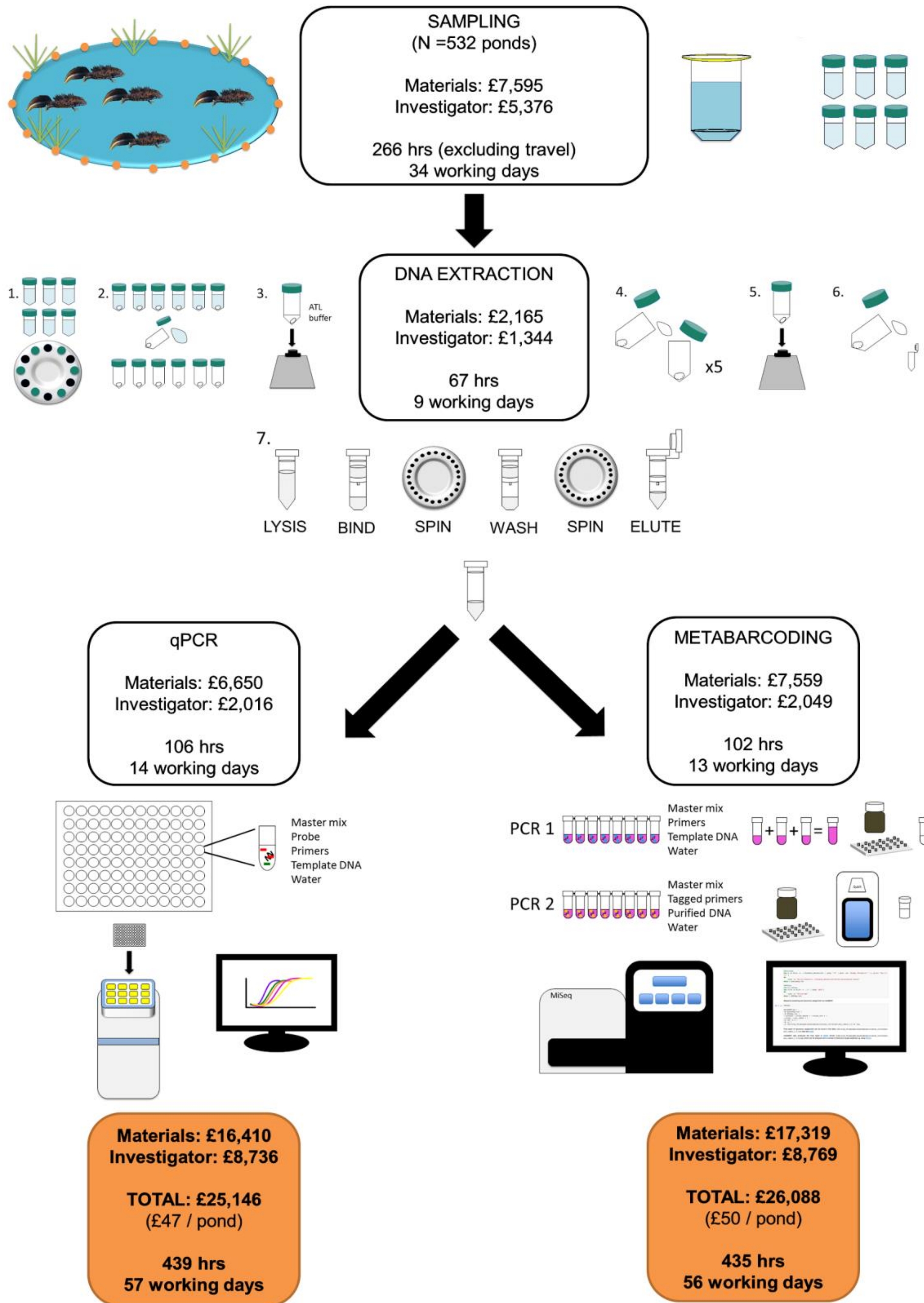


Figure 2.5: Cost and investigator effort required for targeted qPCR of *T. cristatus* and eDNA metabarcoding of vertebrate communities from pond water samples.

2.4 Discussion

We have demonstrated eDNA metabarcoding is a highly sensitive tool for monitoring *T. cristatus* alongside the wider biological community, corroborating other comparisons of eDNA metabarcoding and qPCR for single-species monitoring (Lacoursière-Roussel *et al.*, 2016a; Schneider *et al.*, 2016). Despite reduction in single-species detection, eDNA metabarcoding revealed a wealth of biodiversity information and could enable more effective freshwater monitoring networks and better understanding of community structure and ecosystem function alongside *T. cristatus* monitoring (Biggs *et al.*, 2016). However, both eDNA approaches have advantages and drawbacks which must be considered for design and implementation of biodiversity monitoring programmes.

2.4.1 Single-species detection by qPCR and metabarcoding

A direct comparison of sensitivity between qPCR and metabarcoding is not straightforward: stochasticity in qPCR largely occurs during amplification (volume of template DNA and technical replication), whereas stochastic variation during metabarcoding arises through PCR amplification and sequencing (depth and replication) (Kelly *et al.*, 2014; Thomsen *et al.*, 2016; Deiner *et al.*, 2017). In our study, 12 independent qPCR replicates were performed for each sample but due to limited resources, metabarcoding was based on three pooled PCR replicates which were sequenced once only. Therefore, to enable a fair comparison between methods in terms of PCR effort, a threshold of $\geq$ 4/12 positive replicates (qPCR TA) was applied to the qPCR data. Detection sensitivity was most similar between methods with the qPCR threshold and without the metabarcoding threshold. Both eDNA metabarcoding and qPCR displayed reduced *T. cristatus* detection when thresholds were applied; although, this may reflect reduced false positive detections rather than decreased sensitivity. Lower sensitivity of the eDNA metabarcoding approach used here may also stem from sample degradation during long-term storage. The samples used were stored for more than 12 months at -80 °C before metabarcoding. However, long-term storage and continual freeze-thawing of samples may allow aggregation of inhibitory substances which impair PCR amplification and cause false negatives (Takahara, Minamoto & Doi, 2015).

Despite lower sensitivity, strength of the eDNA signal produced by metabarcoding was correlated with that of qPCR, where both *T. cristatus* average read

count and read proportion broadly increased with qPCR score of eDNA samples. The correlation was inconsistent though, where high average or proportional *T. cristatus* read count did not always correspond to high qPCR score. Biggs *et al*. (2015) also found a variable positive association between qPCR and *T. cristatus* counts, where high qPCR score did not always correlate with high counts. Quantitative data on eDNA concentration are needed to examine the performance of each eDNA approach in relation to the amount of eDNA present, and whether these tools can reliably estimate species abundance. This data can be obtained with highly sensitive qPCR assays, and inclusion of internal DNA standards in sequencing runs for metabarcoding (Ushio *et al.*, 2018a). Nonetheless, our results suggest performance of metabarcoding and qPCR are linked and influenced by external factors. Evans *et al.* (2016) suggested the relative abundance and biomass of a species interact to exert a combined effect on eDNA production rate and subsequent metabarcoding detection. The abundance, biomass, and distribution of *T. cristatus* (Biggs *et al.*, 2015), as well as shedding rate, environmental factors, and eDNA transport (Goldberg *et al.*, 2016; Buxton *et al.*, 2017b), may all influence detection and concentration of eDNA, and inferences made using qPCR and metabarcoding.

The comparison between qPCR and metabarcoding must also be examined in context of the sequencing effort. Here, we sequenced a large number of samples (380 including PCR controls) per run to provide a realistic cost scenario for routine monitoring. Yet, metabarcoding sensitivity would likely improve with an increase in read depth per sample (Kelly *et al.*, 2014). In order to directly compare eDNA signal production by these approaches, it may be necessary to perform sequencing replicates to verify true positives where rare species are expected and generate an “eDNA metabarcoding score” system similar to qPCR (Brandon-Mong *et al.*, 2015; Port *et al.*, 2016; Civade *et al.*, 2016; Thomsen *et al.*, 2016). PCR and sequencing replication in metabarcoding may enhance species detection probability through improved amplification of low abundance or highly degraded DNA (Ficetola *et al.*, 2015; Port *et al.*, 2016) that is readily amplified by qPCR (Lacoursière-Roussel *et al.*, 2016a).

Similarly, sequencing of independent biological replicates, opposed to pseudoreplicates from a single water sample, may improve detection and minimise false negatives produced by eDNA metabarcoding (Andruszkiewicz *et al.*, 2017; Bálint *et al.*, 2018). Currently, 90 mL (6 x 15 mL sampled from 600 mL) water is sampled during *T. cristatus* eDNA survey, followed by ethanol precipitation (Biggs *et al.*, 2015). Whilst this may be appropriate for highly-sensitive targeted qPCR, larger water volumes and

filtration may be required to capture eDNA from less abundant vertebrates and characterise community diversity (Shaw *et al.*, 2016b). Additionally, eDNA from different species, and individuals within species, can be unevenly distributed throughout water bodies and may be concentrated in particular areas (Biggs *et al.*, 2015; Hänfling *et al.*, 2016; Evans *et al.*, 2017a), thus sampling strategies must be carefully designed to ensure eDNA samples are representative of biodiversity present.

Metabarcoding assays are also susceptible to problems from taxon bias, DNA swamping, and bioinformatics related problems (Taberlet *et al.*, 2012; Kelly *et al.*, 2014; Shaw *et al.*, 2016b). Potential reduction in sensitivity of passive community sequencing versus targeted qPCR may relate to the performance of metabarcoding primers for target species. During metabarcoding, DNA from rare species may be masked by highly abundant species (Schneider *et al.*, 2016), or under-represented due to disproportionate eDNA shedding rates across species and preferential amplification of other species (Kelly *et al.*, 2014). PCR-free workflows (i.e. shotgun sequencing) eliminate this bias through indiscriminate sequencing; however, this is unsuitable for conservation projects with target species as a mass of uninformative data are produced, and too costly for routine monitoring schemes (Shaw *et al.*, 2016b; Valentini *et al.*, 2016). We found *T. cristatus* read proportion was negatively associated with post-PCR concentration of eDNA samples. As a positive relationship was not observed, this would suggest PCR amplification with our selected marker and primers was not biased toward our focal species. However, we cannot conclude that our metabarcoding assay was free of primer bias as post-PCR concentration of eDNA samples can be influenced by PCR stochasticity.

Multiple markers (e.g. COI, CytB, 12S, 16S) are increasingly used in eDNA metabarcoding to cast a wider net of species detection and minimise primer bias (Evans *et al.*, 2016b, 2017a; Valentini *et al.*, 2016; Hänfling *et al.*, 2016; Shaw *et al.*, 2016a). Using markers from both mitochondrial and nuclear genes may reduce bias associated with specific genes or primers, and provide greater taxonomic resolution (Kelly *et al.*, 2014). Furthermore, multiple markers of different lengths may enhance understanding of eDNA persistence and state, and species location. Long barcodes bind to stable DNA that has been recently deposited by species (Hänfling *et al.*, 2016), and may reduce false negatives whilst increasing taxonomic resolution and accuracy (Kelly *et al.*, 2014; Valentini *et al.*, 2016; Shaw *et al.*, 2016a). In contrast, short barcodes (such as 12S used here) challenge sequencers and bioinformatics tools (Taberlet *et al.*, 2012; Shaw *et al.*, 2016a), but readily amplify short, degraded DNA fragments that persist longer and

possibly disperse further in water bodies, improving probability of detection (Hänfling *et al.*, 2016). It is possible that metabarcoding detection rates could be improved by using group-specific metabarcoding primers for amphibians, such as the 'batra' set recently designed by Valentini *et al*. (2016). More specific primers could increase relative coverage of *T. cristatus,* providing more comparable detection rates to qPCR. This is worth investigating, but with the caveat that group-specific primers obviously restrict the biodiversity information that can be gained from an ecosystem.

2.4.2 False negatives

This study did not aim to evaluate the sensitivity of molecular methods against standard *T. cristatus* survey methodologies. Egg searches were used to detect false negatives produced by qPCR and metabarcoding and in doing so, revealed some interesting results. Biggs *et al.* (2015) previously demonstrated qPCR had higher detection rate than egg searches (as well as torchlight, netting, and bottle trapping), but here we show this also holds true for metabarcoding. Importantly, absence of eggs does not infer absence of adults or larvae, and this method is highly dependent on weather conditions and water clarity (Rees *et al.*, 2014a; Biggs *et al.*, 2015). Despite considerably higher detection rate of both eDNA approaches, eggs were recorded in a small number of ponds that were eDNA negative. eDNA analysis can incorrectly infer absence or low abundance of species if inhibition or interference from non-target DNA has occurred (Goldberg *et al.*, 2016). Alternatively, eDNA false negatives may have been a by-product of sampling strategy and effort for *T. cristatus*. Larger water volumes and/or more biological replication instead of pseudoreplication (established *T. cristatus* eDNA sampling strategy) may improve detection (Lopes *et al.*, 2016; Andruszkiewicz *et al.*, 2017; Bálint *et al.*, 2018). All methods revealed *T. cristatus* in ponds where other approaches failed, emphasising that these species monitoring tools are complementary and should be used in combination to achieve maximum detection probability. However, integrative strategies combining molecular and conventional tools are often not cost-efficient for most applications.

2.4.3 False positives

False positives may arise from field contamination and eDNA transport in the environment - particularly by waterfowl (Shaw *et al.*, 2016a). eDNA is retained by

predators, discarded in faeces, and transported by anthropogenic activity, combined with natural water currents and flow (Hänfling *et al.*, 2016). In the laboratory, PCR-accumulated and sequencing error, including primer mismatch (Andersen *et al.*, 2012) and 'tag jumps' (Schnell, Bohmann & Gilbert, 2015), can induce misassignment leading to false positives, cross-contamination between samples, or laboratory contamination (Andruszkiewicz *et al.*, 2017).

False positives can be modelled and estimated using site occupancy modelling of metabarcoding data (Ficetola *et al.*, 2015) or risk of false positives minimised using a sequencing threshold, that is the number of sequence reads required for a sample to be species positive (Hänfling *et al.*, 2016; Civade *et al.*, 2016; Evans *et al.*, 2017a). However, such thresholds can reduce detection of rare species, a primary goal of this study, and may fail where false and true positives occur at similar frequency (Hänfling *et al.*, 2016). Instead, we calculated species-specific sequence thresholds to more accurately control for false positives in our dataset without compromising *T. cristatus* detection.

In our study, *H. sapiens* DNA occurred at high frequency and abundance; this may have been a true environmental signal from pond water, or real contaminant as encountered in other metabarcoding research (Port *et al.*, 2016; Valentini *et al.*, 2016; Thomsen *et al.*, 2016). Blocking primers can prevent amplification of abundant non-target DNA like *H. sapiens* (Valentini *et al.*, 2016) but may fail (Thomsen *et al.*, 2016) or prevent amplification of target taxa (Port *et al.*, 2016). Alongside *H. sapiens*, other aquatic and terrestrial vertebrate DNA occurred at high frequency in NTCs, although these were not removed by addition of mineral oil. An even stricter forensic laboratory set-up, such as that employed for ancient DNA (aDNA), should be adopted to ensure data robustness. Positive and negative controls should be included at each stage of metabarcoding workflows to monitor contamination (Deiner *et al.*, 2017). However, preventive measures inevitably increase research cost and some degree of contamination is unavoidable in metabarcoding (Kelly *et al.*, 2014; Brandon-Mong *et al.*, 2015; Port *et al.*, 2016; Thomsen *et al.*, 2016).

Our results also highlight the importance and impact of qPCR thresholds when inferring species presence-absence. Similar to Smart *et al*. (2016), we found a stringent qPCR threshold reduced detection sensitivity. As yet, no guidance exists to indicate how many samples or replicates must be positive to class a site as species-positive (Goldberg *et al.*, 2016; Smart *et al.*, 2016) but clearly this must be addressed to improve standardisation and reproducibility of eDNA research. Importantly, less stringent

thresholds (and false positives inherent to these) are somewhat precautionary and may better protect *T. cristatus* by preventing development. Therefore, whilst reduction or removal of false positives is desirable, detection thresholds must not compromise protection of threatened species either. Until a suitable threshold can be established, it may be more appropriate to re-analyse samples which yield one positive qPCR replicate to prevent false positives (Rees *et al.*, 2014b; Goldberg *et al.*, 2016).

2.4.4 Cost and investigator effort

Cost efficiency combined with the overarching aim of a monitoring or conservation programme should always be considered. We found eDNA metabarcoding was more costly than qPCR but both approaches required similar investigator effort. qPCR scales to the number of samples being processed (Schneider *et al.*, 2016) whereas metabarcoding has fixed costs including reagent kit for high-throughput sequencing platform (Bálint *et al.*, 2018). eDNA metabarcoding becomes more cost-efficient as more samples are processed (Bálint *et al.*, 2018) but fewer replicates would reduce qPCR cost (Davy *et al.*, 2015; Smart *et al.*, 2016). Travel was excluded from our cost estimate but inclusion of this expense would further reduce cost efficiency of both approaches. Cost of eDNA monitoring is influenced by sample size, methods, replication, laboratory, statistical power, and occupancy modelling (Davy *et al.*, 2015; Evans *et al.*, 2017b). Consequently, cost is proportional to project requirements (Davy *et al.*, 2015) and will vary depending on choice of qPCR or metabarcoding workflow. Whilst qPCR is established technology that has reached its price ceiling, high-throughput sequencing is relatively new technology and prices will continue to drop, meaning higher sample throughput and more technical replication will be possible. We therefore argue that metabarcoding will become more cost-efficient in the long-term, providing more data at lower cost and comparable sensitivity to qPCR. However, where samples cannot be processed in large batches, qPCR may retain cost-efficiency.

2.4.5 Conclusion

eDNA metabarcoding holds promise for holistic biodiversity monitoring of freshwater ponds as opposed to targeted qPCR for flagship or indicator species such as *T. cristatus*. Metabarcoding can reveal entire species assemblages from environmental samples

without prior ecosystem information and provide broad-scale distribution data for multiple species simultaneously. Nonetheless, the method at present appears to be less sensitive than qPCR for single-species monitoring, and species detection by molecular and conventional methods was incongruent. Comprehensive study of the influence of water volume, eDNA capture and extraction method, and sample storage on single-species and community detection in lentic and lotic systems is required. Minimising the risk of false positives and contamination remains a pressing issue in metabarcoding, and standard contamination measures (Goldberg *et al.*, 2016) may be insufficient for analysis of vertebrate assemblages. Currently, cost and investigator effort required for metabarcoding and qPCR are broadly equivalent, but reduced sequencing costs may level the playing field. We conclude that eDNA metabarcoding is not yet a replacement for targeted qPCR and conventional survey, but rather another tool in the ecologist toolbox. Ultimately, choice of monitoring tool(s) is specific to the aims of each conservation project. At present, qPCR retains sensitivity for *T. cristatus* populations of all sizes, regardless of sample number processed. Under a realistic conservation monitoring scenario, where funding is limited and samples must be processed in large batches, metabarcoding may suffer from false negatives due to reduced sequencing depth and replication. However, in many cases, the biodiversity information generated by this approach, and its implications for community ecology and conservation, will eclipse lower sensitivity. This passive screening approach would be most effective for initial survey of water bodies to generate broad-scale multi-species distribution data. This holistic data can then inform best use of funding and time for targeted species-specific survey.

2.5 Acknowledgements

This work was funded by the University of Hull. We would like to thank Jennifer Hodgetts (Fera Science Ltd) for assisting with sample collection and data dissemination, and Jianlong Li (University of Hull) for primer design and advice on laboratory protocols. Furthermore, Barbara Mabel, Elizabeth Kilbride (University of Glasgow), Andrew Buxton and Richard Griffiths (DICE, University of Kent) provided tissue samples for primer validation and Sanger sequencing to supplement the reference database.

2.6 Data accessibility

Raw sequence reads have been archived on the NCBI Sequence Read Archive (Bioproject: PRJNA417951; SRA accessions: SRR6285413 - SRR6285678). Jupyter notebooks, R scripts and corresponding data are deposited in a dedicated GitHub repository (https://github.com/HullUni-bioinformatics/Harper_et_al_2018) which has been permanently archived (https://doi.org/10.5281/zenodo.2633978).

Chapter 3: Ground-truthing environmental DNA (eDNA) metabarcoding for ecological hypothesis testing at the pondscape

Example of a pondscape in Alaska, USA
© user: Travis | Flickr | CC BY-NC 2.0

This chapter is available online as

Harper, L.R., Handley, L.L., Hahn, C., Boonham, N., Rees, H.C., Lewis, E., Adams, I.P., Brotherton, P., Phillips, S. & Hänfling, B. (2019) Testing ecological hypotheses at the pondscape with environmental DNA metabarcoding: a case study on a threatened amphibian. *bioRxiv*, 278309. https://doi.org/10.1101/278309

Abstract

Environmental DNA (eDNA) metabarcoding is revolutionising biodiversity monitoring, but has unrealised potential for ecological hypothesis testing. Here, we ground-truth eDNA metabarcoding for describing vertebrate communities from 532 UK ponds. We examine associations between the threatened great crested newt (*Triturus cristatus*), a flagship conservation species, and other vertebrates. Furthermore, we investigate factors influencing *T. cristatus* occurrence and vertebrate species richness at the pondscape. *T. cristatus* occurrence was positively correlated with amphibian and waterfowl species richness, where *T. cristatus* had strong positive associations with smooth newt (*Lissotriton vulgaris*), common coot (*Fulica atra*), and common moorhen (*Gallinula chloropus*), but a negative association with common toad (*Bufo bufo*). *T. cristatus* occurrence was negatively correlated to fish species richness, specifically three-spined stickleback (*Gasterosteus aculeatus*) and ninespine stickleback (*Pungitius pungitius*) presence. Both *T. cristatus* occupancy and vertebrate species richness correlated with the *T. cristatus* Habitat Suitability Index score, supporting its application to *T. cristatus* survey. We reaffirm reported associations (e.g. *T. cristatus* preference for deeper ponds) but also provide novel insights, including a negative effect of pond outflow on *T. cristatus*. Furthermore, we reveal novel factors influencing vertebrate species richness at the pondscape, including pond density, macrophyte cover, and terrestrial habitat. Our findings demonstrate the prospects of eDNA metabarcoding for hypothesis testing at landscape scale and dramatic enhancement of freshwater conservation, management, monitoring and research.

3.1 Introduction

Environmental DNA (eDNA) analysis offers ecologists exceptional power to detect organisms within and across ecosystems. DNA released by organisms into their environment via secretions, excretions, gametes, blood, or decomposition, can be sampled and analysed using different approaches to reveal the distribution of single or multiple species (Rees *et al.*, 2014b; Lawson Handley, 2015). When combined with high-throughput sequencing (i.e. eDNA metabarcoding), eDNA can yield efficient, comprehensive assessments of entire communities (Deiner *et al.*, 2017), providing a step change in biodiversity monitoring (Hering *et al.*, 2018). eDNA metabarcoding has untapped potential to test ecological hypotheses by enabling biodiversity monitoring at landscape scales with minimal impact to communities under investigation. Although this tool has been used to estimate species richness and assess diversity along environmental gradients (e.g. Hänfling *et al.*, 2016; Olds *et al.*, 2016; Kelly *et al.*, 2016; Evans *et al.*, 2017a), its applications in community ecology are relatively unexplored.

Aquatic ecosystems are highly suited to eDNA studies (Muha *et al.*, 2017) as eDNA exists in multiple states with rapid modes of transport and degradation, increasing detectability of contemporary biodiversity (Rees *et al.*, 2014b; Barnes & Turner, 2015). Lentic systems provide further opportunities for eDNA research, being discrete water bodies with variable physicochemical properties that do not experience flow dynamics (Muha *et al.*, 2017). Ponds in particular have enormous biodiversity and experimental virtue that has not been maximised in previous eDNA metabarcoding assessments of this habitat (Valentini *et al.*, 2016; Evans *et al.*, 2017a; Klymus *et al.*, 2017b; Ushio *et al.*, 2017; Bálint *et al.*, 2018). These small and abundant water bodies span broad ecological gradients (De Meester *et al.*, 2005) and comprise pondscapes - a network of ponds and their surrounding terrestrial habitat (Hill *et al.*, 2018). Pondscapes contribute substantially to aquatic and non-aquatic biodiversity across spatial scales, with ponds supporting many rare and protected species in fragmented landscapes (De Meester *et al.*, 2005; Biggs *et al.*, 2016; Hill *et al.*, 2018). Consequently, ponds are model systems for experimental validation and examination of biogeographical patterns (De Meester *et al.*, 2005). Habitat complexity and tools required for different taxa with associated bias (Evans *et al.*, 2017a) and cost (Valentini *et al.*, 2016) once hindered exhaustive sampling of pond biodiversity (Hill *et al.*, 2018), but eDNA metabarcoding may overcome these barriers.

In the UK, the threatened great crested newt (*Triturus cristatus*) is a flagship

species for pond conservation. The extensive literature on *T. cristatus* provides an excellent opportunity to ground truth ecological patterns revealed by eDNA metabarcoding. Both biotic (e.g. breeding substrate, prey, and predators) and abiotic (e.g. pond area, depth, and temperature) factors are known to influence *T. cristatus* breeding success (Langton, Beckett & Foster, 2001). The *T. cristatus* Habitat Suitability Index (HSI; Oldham *et al.*, 2000; ARG-UK, 2010) accounts for these factors using 10 suitability indices that are scored and combined to calculate a decimal score between 0 and 1 (where 1 = excellent habitat). Larvae are susceptible to fish and waterfowl predation (Rannap & Briggs, 2006; Skei *et al.*, 2006; Hartel, Nemes & Oellerer, 2010), and adults reportedly avoid ponds containing three-spined stickleback (*Gasterosteus aculeatus*) (McLee & Scaife, 1992), ninespine stickleback (*Pungitius pungitius*), crucian carp (*Carassius carassius*), and common carp (*Carassius carpio*) (Rannap, Lõhmus & Briggs, 2009a b). Conversely, *T. cristatus* and the smooth newt (*Lissotriton vulgaris*) prefer similar habitat and often co-occur (Rannap & Briggs, 2006; Skei *et al.*, 2006; Rannap *et al.*, 2009a; Denoël *et al.*, 2013). *T. cristatus* individuals thrive in ponds with good water quality as indicated by diverse macroinvertebrate communities (Oldham *et al.*, 2000; Rannap *et al.*, 2009a), and water clarity is important for breeding displays, foraging success, and egg survival (Rannap & Briggs, 2006; Skei *et al.*, 2006). Pond networks encourage *T. cristatus* occupancy (Joly *et al.*, 2001; Rannap *et al.*, 2009a; Hartel *et al.*, 2010; Denoël *et al.*, 2013), but larger area discourages presence (Joly *et al.*, 2001). Ponds with heavy shading (Vuorio, Heikkinen & Tikkanen, 2013) or dense macrophyte cover (Rannap & Briggs, 2006; Skei *et al.*, 2006; Hartel *et al.*, 2010) are unlikely to support viable populations. *T. cristatus* individuals also depend on terrestrial habitat, preferring open, semi-rural pondscapes (Denoël *et al.*, 2013) containing pasture, extensively grazed and rough grassland, scrub, and coniferous and deciduous woodland (Oldham *et al.*, 2000; Rannap & Briggs, 2006; Rannap *et al.*, 2009a; Gustafson, Malmgren & Mikusiński, 2011; Vuorio *et al.*, 2013).

We assessed vertebrate communities at the pondscape using a dataset generated by eDNA metabarcoding for over 500 ponds with comprehensive environmental metadata. We evaluated eDNA metabarcoding as a tool for ecological hypothesis testing, and compared its outputs to previous results generated by established methods. Specifically, we aimed to identify biotic (community presence-absence data) and abiotic determinants (environmental metadata on ponds and surrounding terrestrial habitat) of *T. cristatus* at an unparalleled scale, and determinants of vertebrate species richness at the

pondscape - an impractical task by conventional means. Finally, we investigated applicability of the HSI to predict eDNA-based *T. cristatus* occupancy and vertebrate species richness of ponds.

3.2 Materials and methods

3.2.1 Samples

We repurposed the taxonomically assigned sequence reads from Chapter 2 that were produced using eDNA metabarcoding of pond water to compare quantitative PCR and eDNA metabarcoding for *T. cristatus* detection (see also Harper *et al*., 2018a). Samples from 508 ponds included in Natural England's Great Crested Newt Evidence Enhancement Programme were processed using eDNA metabarcoding alongside 24 privately surveyed ponds. Water samples were collected using established methodology (Biggs *et al.*, 2015), detailed in Appendix 2.1. In brief, 20 x 30 mL water samples were collected from each pond and pooled. Six 15 mL subsamples were taken from the pooled sample and each added to 33.5 mL absolute ethanol and 1.5 mL sodium acetate 3 M (pH 5.2). Subsamples were pooled during DNA extraction to produce one eDNA sample per pond. Targeted quantitative PCR detected *T. cristatus* in 265 (49.81%) ponds (see Chapter 2 and Harper *et al.*, 2018a).

Environmental metadata (Table S3.1) were collected for 504 of 532 ponds (Fig. S3.1) by environmental consultants contracted for Natural England's Great Crested Newt Evidence Enhancement Programme. Metadata included: maximum depth; circumference; width; length; area; density (i.e. number of ponds per km^2); terrestrial overhang; shading; macrophyte cover; HSI score (Oldham *et al.*, 2000); HSI band (categorical classification of HSI score from ARG-UK, 2010); permanence; water quality; pond substrate; presence of inflow or outflow; presence of pollution; presence of other amphibians, fish and waterfowl; woodland; rough grass; scrub/hedge; ruderals; other good terrestrial habitat (i.e. good terrestrial habitat that did not conform to aforementioned habitat types); and overall terrestrial habitat quality.

3.2.2 DNA reference database construction

A custom, phylogenetically curated reference database of mitochondrial 12S ribosomal RNA (rRNA) sequences for UK fish species was previously constructed for eDNA metabarcoding of lake fish communities (Hänfling *et al.*, 2016). In Chapter 2, additional reference databases for UK amphibians, reptiles, birds, and mammals were constructed (see Harper *et al*. 2018a and Appendix 2.1). Reference sequences available for species varied across vertebrate groups: amphibians 100.00% ($N = 21$), reptiles 90.00% ($N = 20$), mammals 83.93% ($N = 112$), and birds 55.88% ($N = 621$). Table S3.2 lists species without database representation, i.e. no records for any species in a genus. Sanger sequences were obtained from tissue of *T. cristatus*, *L. vulgaris*, Alpine newt (*Ichthyosaura alpestris*), common toad (*Bufo bufo*), and common frog (*Rana temporaria*) to supplement the amphibian database (see Appendix 2.1). The complete reference databases compiled in GenBank format have been deposited in a dedicated GitHub repository for Chapter 2, permanently archived at: https://doi.org/10.5281/zenodo.2633978.

3.2.3 Primer validation

Reference databases were combined for *in silico* validation of published 12S rRNA primers 12S-V5-F (5'-ACTGGGATTAGATACCCC-3') and 12S-V5-R (5'-TAGAACAGGCTCCTCTAG-3') (Riaz *et al.*, 2011) using ecoPCR software (Ficetola *et al.*, 2010). Set parameters allowed a 50-250 bp fragment and three mismatches between each primer and reference sequence. Primers were validated *in vitro* for UK fish by Hänfling *et al*. (2016) and here for six UK amphibian species (Fig. S3.2).

3.2.4 eDNA metabarcoding

We used the taxonomically assigned sequence reads generated using vertebrate eDNA metabarcoding in Chapter 2 and Harper *et al*. (2018a). The eDNA metabarcoding workflow is fully described in Appendix 2.1 and Harper *et al*. (2018a). eDNA was first amplified with the aforementioned primers, where PCR positive controls (six per PCR plate; $n = 114$) were cichlid (*Rhamphochromis esox*) DNA (0.284 ng/µL) and PCR negative controls (six per PCR plate; $n = 114$) were sterile molecular grade water (Fisher Scientific UK Ltd, UK). PCR products were individually purified using E.Z.N.A® Cycle

Pure V-Spin Clean-Up Kits (Omega Bio-tek, GA, USA) following the manufacturer's protocol. The second PCR bound Multiplex Identification (MID) tags to the purified products. PCR products were individually purified using magnetic bead clean-up and quantified with a Quant-IT™ PicoGreen™ dsDNA Assay (Invitrogen, UK). Samples were normalised, pooled, and libraries quantified using a Qubit™ dsDNA HS Assay (Invitrogen, UK). Libraries were sequenced on an Illumina MiSeq using 2 x 300 bp V3 chemistry (Illumina, Inc, CA, USA) and raw sequence reads processed using metaBEAT (metaBarcoding and Environmental Analysis Tool) v0.8 (https://github.com/HullUni-bioinformatics/metaBEAT). After quality trimming, merging, chimera detection, and clustering, non-redundant query sequences were compared against our reference database using BLAST (Zhang *et al.*, 2000). Putative taxonomic identity was assigned using a lowest common ancestor (LCA) approach based on the top 10% BLAST matches for any query matching with at least 98% identity to a reference sequence across more than 80% of its length. Unassigned sequences were subjected to a separate BLAST against the complete NCBI nucleotide (nt) database at 98% identity to determine the source via LCA as described above. The bioinformatic analysis has been deposited in the Chapter 2 GitHub repository for reproducibility.

3.2.5 Data analysis

Analyses were performed in the statistical programming environment R v.3.4.2 (R Core Team 2017). Data and R scripts have been deposited in a dedicated GitHub repository for this chapter, permanently archived at: https://doi.org/10.5281/zenodo.2634033. Assignments from different databases were merged, and spurious assignments (i.e. non-UK species, invertebrates and bacteria) removed from the dataset. The family Cichlidae was reassigned to *Rhamphochromis esox*. The green-winged teal (*Anas carolinenisis*) was reassigned to *Anas* (Dabbling ducks) because this species is a rare migrant and reference sequences were identical to those for mallard (*Anas platyrhynchos*) and Eurasian teal (*Anas crecca*), which are widely distributed across the UK. Scottish wildcat (*Felis silvestris*) does not occur at the sampling localities (Kent, Lincolnshire and Cheshire) and was therefore reassigned to domestic cat (*Felis catus*). Wild boar (*Sus scrofa*) and grey wolf (*Canis lupus*) were reassigned to domestic pig (*Sus scrofa domesticus*) and domestic dog (*Canis lupus familiaris*) given the restricted distribution of *S. scrofa* and absence of *C. lupus* in the UK. The genus *Strix* was reassigned to tawny owl (*Strix aluco*) as it is the

only UK representative of this genus. Where family and genera assignments containing a single UK representative had reads assigned to species, reads from all assignment levels were merged and manually assigned to that species. Higher taxonomic assignments excluding the genus *Anas* were then removed, thus taxonomic assignments in the final dataset were predominantly of species resolution.

To minimise risk of false positives, species were only classed as present at sites if their sequence frequency exceeded species-specific thresholds. Thresholds were defined using the maximum sequence frequency of each species in PCR positive controls (n = 114; Table S3.3). For example, the *T. cristatus* threshold was 0.028% to omit false positives in PCR positive controls. After applying thresholds, the read count data were converted to a species presence-absence matrix. Analyses were based on species-specific thresholds, but also performed for different blanket sequence thresholds (0.05 - 30%, Tables S3.4-3.9). We tested biotic and abiotic determinants of *T. cristatus* occupancy and vertebrate species richness, and appropriateness of the HSI. Hypotheses are summarised in Table 3.1.

All Generalized Linear Mixed-effects Models (GLMMs) were executed using the R package lme4 v1.1-12 (Bates *et al.*, 2015). *T. cristatus* occurrence relating to number of other vertebrate species was investigated with a binomial GLMM, and species associations identified using the R package cooccur v1.3 (Griffith, Veech & Marsh, 2016) (N = 532). Identified associations informed candidate biotic variables for inclusion with abiotic variables (Table S3.1) in a binomial GLMM of *T. cristatus* occurrence (n = 504). Candidate explanatory variables were assessed for collinearity, relative importance, and non-linearity (see Appendix 3.1). HSI score and band were collinear, thus HSI score was analysed in a separate binomial GLMM. Using the R package ncf v1.1-7 (Bjørnstad, 2017), spline correlograms of the Pearson residuals from a binomial Generalized Linear Model (GLM) and GLMM were compared to assess spatial autocorrelation potential in our dataset. Sample was treated as a random effect in the GLMM to account for spatial dependencies between ponds (Zuur *et al.*, 2009). The mixed model successfully accounted for spatial autocorrelation within sites. The same steps were performed to identify explanatory variables and a modeling framework for vertebrate species richness. A Poisson distribution was specified for all species richness models as the response variable was integer count data.

Binomial and Poisson models considered respectively were nested thus best models were chosen using stepwise backward deletion of terms based on Likelihood

Ratio Tests (LRTs). Akaike's Information Criteria (AIC) was employed to select the most parsimonious model (Akaike, 1973). Final models were tested for overdispersion using the R package RVAideMemoire v0.9-45-2 (Hervé, 2015) and a custom function testing overdispersion of the Pearson residuals. Model fit was assessed using the Hosmer and Lemeshow Goodness of Fit Test (Hosmer & Lemeshow, 2000) within the R package ResourceSelection v0.2-4 (Lele, Keim & Solymos, 2016), quantile-quantile plots, and partial residual plots (Zuur *et al.*, 2009). Model predictions were obtained using the *predictSE* function in the AICcmodavg package v2.0-3 (Mazerolle, 2016) and upper and lower 95% CIs were calculated from the standard error of the predictions. Results were plotted using the R package ggplot2 v2.1.0 (Wickham, 2016).

3.3 Results

3.3.1 eDNA metabarcoding

A total of 532 eDNA samples and 228 PCR controls were processed across two sequencing runs. The runs generated raw sequence read counts of 36,236,862 and 32,900,914 respectively. After trimming and merging of paired-end reads, 26,294,906 and 26,451,564 sequences remained. Following removal of chimeras and redundancy via clustering, the libraries contained 14,141,237 and 14,081,939 sequences (average read counts of 36,826 and 36,671 per sample respectively), of which 13,126,148 and 13,113,143 sequences were taxonomically assigned. The final dataset (thresholds applied and assignments corrected) contained 60 vertebrate species (Table S3.10), including six amphibians, 14 fish, 17 birds, and 22 mammals (Fig. S3.3).

3.3.2 Pondscape biodiversity

All native amphibians were found as well as the non-native marsh frog (*Pelophylax ridibundus*). *T. cristatus* (*n* = 149), *L. vulgaris* (*n* = 152) and *R. temporaria* (*n* = 120) were widespread, but *B. bufo* (*n* = 42), palmate newt (*Lissotriton helveticus*, *n* = 5) and *P. ridibundus* were uncommon (*n* = 1). The threatened European eel (*Anguilla anguilla*, *n* = 15), European bullhead (*Cottus gobio*, *n* = 14), and *C. carassius* (*n* = 2) were detected alongside native fishes, such as pike (*Esox lucius*, *n* = 17) and roach (*Rutilus rutilus*, *n* =

72), but also introduced species, including *C. carpio* (n = 41), ruffe (*Gymnocephalus cernua*, n = 1), and rainbow trout (*Oncorhynchus mykiss*, n = 3). Waterfowl identified ranged from common moorhen (*Gallinula chloropus*, n = 215) to grey heron (*Ardea cinerea*, n = 1) and Eurasian oystercatcher (*Haematopus ostralegus*, n = 1). Terrestrial fauna were often detected in fewer than five ponds (Fig. S3.3c, d). Buzzard (*Buteo buteo*, n = 4), Eurasian jay (*Garrulus glandarius*, n = 7), dunnock (*Prunella modularis*, n = 4), and starling (*Sturnus vulgaris*, n = 4) were the most frequently detected terrestrial birds. Domesticated, including cow (*Bos taurus*, n = 179) and pig (*Sus scrofa domesticus*, n = 140), or introduced mammals (Mathews *et al.*, 2018), such as grey squirrel (*Sciurus carolinensis*, n = 57) and Reeve's muntjac (*Muntiacus reevesi*, n = 3), outweighed native mammals. Nonetheless, we detected several mammals with Biodiversity Actions Plans and/or of conservation concern (Mathews *et al.*, 2018), including otter (*Lutra lutra*, n = 1), water vole (*Arvicola amphibius*, n = 16), European polecat (*Mustela putorius*, n = 1), brown hare (*Lepus europaeus*, n = 1) and water shrew (*Neomys fodiens*, n = 9). Notably, the invasive American mink (*Neovison vison*) was absent despite widespread UK distribution (Mathews *et al.*, 2018). All species and their detection frequencies are listed in Table S3.10.

3.3.3 Biotic determinants of *T. cristatus* occurrence

T. cristatus occupancy was negatively correlated with fish species richness, but positively influenced by amphibian and waterfowl species richness (Fig. 3.1, GLMM: overdispersion χ^2_{525} = 517.636, P = 0.582; fit χ^2_{8} = 22.524, P = 0.004, R^2 = 9.43%). *T. cristatus* had significant (P < 0.05) positive associations with four species (Fig. 3.2), including *L. vulgaris*, common coot (*Fulica atra*), *G. chloropus*, and *S. s. domesticus*. However, *T. cristatus* had significant (P < 0.05) negative associations with five species (Fig. 3.2), including *B. bufo*, *P. pungitius*, *G. aculeatus*, *S. carolinensis*, and common pheasant (*Phasianus colchicus*). Only presence-absence of *L. vulgaris*, *B. bufo*, *G. aculeatus*, and *S. carolinensis* were retained by model selection as explanatory variables for the GLMM (Figs. 3.3a-d) with abiotic determinants. Results of analyses are summarised and compared to previously reported determinants in Table 3.1.

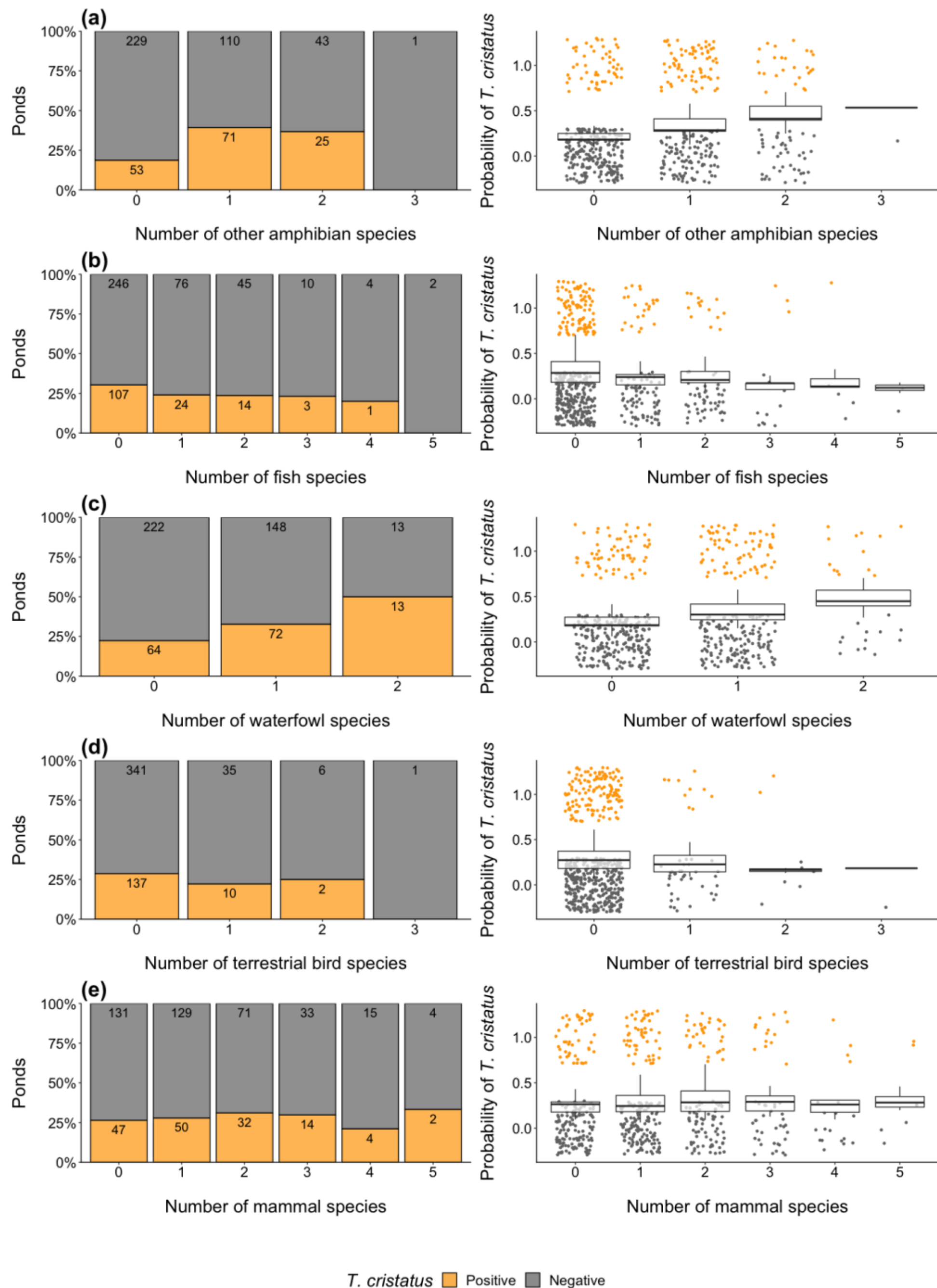


Figure 3.1: ***T. cristatus*** **presence (orange) and absence (grey) in relation to number of species from different vertebrate groups detected by eDNA** (N = 532 ponds): **(a)** other amphibians, **(b)** fish, **(c)** waterfowl, **(d)** terrestrial birds, and **(e)** mammals. Observed proportion of ponds with and without *T. cristatus* (left) is plotted alongside predicted probability of *T. cristatus* occurrence in ponds as determined by the binomial GLMM (right). Numbers on barplots of observed occupancy are the number of ponds for each category. In plots showing predicted *T. cristatus* occupancy, the observed data is shown as points which have been jittered around 0 and 1 to clarify variation in point density. Boxes are the model predictions.

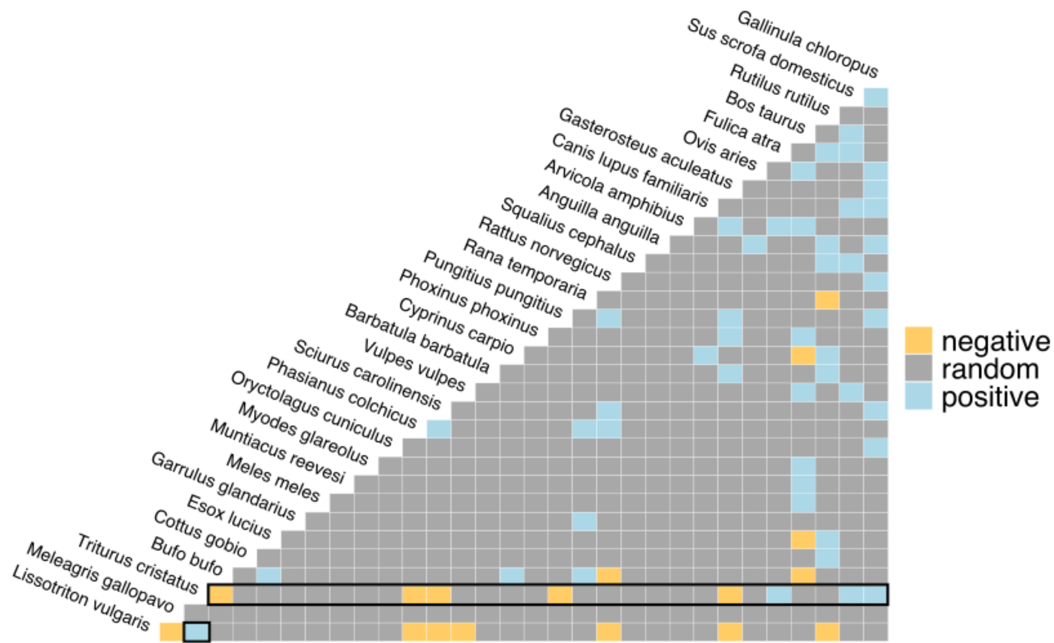


Figure 3.2: Heat map showing significant ($P < 0.05$) positive and negative species associations determined by the probabilistic co-occurrence model for the eDNA metabarcoding presence-absence data ($N = 532$ ponds). Species names are positioned to indicate the columns and rows that represent their pairwise relationships with other species. Species are ordered by those with the most negative interactions to those with the most positive interactions (left to right). Associations relevant to *T. cristatus* are highlighted in black.

Table 3.1: Summary of established and novel abiotic and biotic determinants of *T. cristatus* occupancy. Reported effects on *T. cristatus* occupancy in the literature and hypothesised effects on eDNA-based *T. cristatus* occurrence are given for each determinant. Any determinants not reported in the literature are listed as UNK. Direction of observed effects on eDNA-based *T. cristatus* occupancy determined by each analysis (GLMM assessing number of species in each vertebrate group, $N = 532$; co-occur analysis, $N = 532$; GLMM combining abiotic and biotic factors $n = 504$; and GLMM assessing HSI, $n = 504$) are given. No, negative and positive effects are listed as 0, - and + respectively. For categorical variables with more than one level, effect size and standard error (SE) are only given for levels reported in the model summary. Test statistic is for LRT used and significant P-values (<0.05) are in bold. Variables included for model selection but not retained in the final model are listed as NR. Co-occur analysis was not applicable (NA) to abiotic factors.

Determinant	**Effect reported**	**Hypothesised effect**	**Analysis**					
			Cooccur		**GLMM**			
			Effect	***P***	**DF**	**Effect size (SE)**	**χ^2**	***P***
Fish	-/0	-			1	-0.239 (0.124)	4.065	**0.044**
G. aculeatus	-	-	-	**0.009**	1	-1.432 (0.561)	9.453	**0.002**
P. pungitius	-	-	-	**0.047**				
C. carpio	-	-						
C. carassius	-	-						
Waterfowl	-	-			1	0.617 (0.181)	13.050	**<0.001**
F. atra	UNK		+	**0.023**				
G. chloropus	UNK		+	**<0.001**				
Amphibians	UNK				1	0.558 (0.149)	16.641	**<0.001**
L. vulgaris	+	+	+	**< 0.001**	1	1.081 (0.303)	17.434	**<0.001**
B. bufo	UNK		-	**0.009**	1	-1.635 (0.696)	8.228	**0.004**
Terrestrial birds	UNK				1	-0.335 (0.291)	1.444	0.230
P. colchicus	UNK		-	**0.048**				
Terrestrial mammals	UNK				1	0.028 (0.091)	0.095	0.758
S. carolinensis	UNK		-	**0.018**	1	-1.591 (0.534)	12.432	**<0.001**
S. s. domesticus	UNK		+	**0.004**				
Pond area	-/+	-	NA	NA	1	-0.0004 (0.0002)	6.453	**0.011**

Pond density	+	+	NA	NA		NR		
Pond depth	+	+	NA	NA	1	0.282 (0.139)	4.266	**0.039**
Water quality	+	+	NA	NA		NR		
Outflow	UNK		NA	NA	1	-0.713 (0.359)	4.467	**0.035**
Macrophyte cover	-/+	-	NA	NA		NR		
Shading	-/+	-	NA	NA		NR		
Woodland	+	+	NA	NA		NR		
Grassland	+	+	NA	NA		NR		
HSI	0/+	+	NA	NA	1	3.020 (0.791)	15.709	**<0.001**
Ruderal	UNK		NA	NA	2		6.507	**0.039**
None						-0.617 (0.527)		
Some						0.032 (0.528)		
Other good terrestrial habitat	UNK		NA	NA	2		7.918	**0.019**
None						0.428 (0.429)		
Some						-0.316 (0.424)		
Species richness	UNK		NA	NA	1	0.527 (0.105)	60.267	**<0.001**

T. cristatus individuals were more likely to occupy ponds with more amphibian species (Fig. 3.1a). *T. cristatus* was detected in 51.97% of ponds (n = 152) containing *L. vulgaris*, but in only 11.91% of ponds (n = 42) with *B. bufo* (Fig. S3.3a). *T. cristatus* occurrence probability was lower in ponds with more fish species, and *T. cristatus* was absent from ponds with more than four fish species (Fig. 3.1b). *T. cristatus* was only found in 14.29% (n = 56) and 6.67% (n = 15) of ponds inhabited by *G. aculeatus* and *P. pungitius* respectively (Fig. S3.3b). In contrast, *T. cristatus* individuals were more likely to occur in ponds with more waterfowl species (Fig. 3.1c). *T. cristatus* occupied 41.67% (n = 48) and 35.81% (n = 215) of ponds with *F. atra* and *G. chloropus* respectively (Fig. S3.3c).

No significant relationships between terrestrial bird or mammal species richness and *T. cristatus* occupancy were found (Figs. 3.1d, e). Yet, species associations between *T. cristatus* and *S. s. domesticus*, *P. colchicus*, and *S. carolinensis* were identified. *T. cristatus* was detected in 37.14% of ponds (n = 140) where *S. s. domesticus* was present (Fig. S3.3d) as opposed to 12.00% (n = 25) and 15.79% (n = 57) of ponds with *P. colchicus* (Fig. S3.3c) and *S. carolinensis* (Fig. S3.3d) records respectively. *T. cristatus* occupancy positively correlated with overall vertebrate species richness (Fig. 3.3h), irrespective of individual species associations.

3.3.4 Abiotic determinants of *T. cristatus* occurrence

Five abiotic determinants were retained in the GLMM explaining *T. cristatus* occupancy (GLMM: overdispersion χ^2_{490} = 413.394, P = 0.995; fit χ^2_8 = 11.794, P = 0.161, R^2 = 38.58%). The probability of *T. cristatus* occurrence increased with greater pond depth but decreased in ponds with larger area, outflow, some other good terrestrial habitat, and where ruderal habitat was absent (Table 3.1, Figs. 3.3e-g, i-j).

3.3.5 Abiotic determinants of vertebrate species richness

Our analysis (Table 3.2, GLMM: overdispersion χ^2_{494} = 431.959, P = 0.979; fit χ^2_8 = -42.708, P = 1.000, R^2 = 8.94%) revealed species richness was greater in ponds with outflow (Fig. 3.4a), but reduced in those with some rough grass habitat compared to ponds with no or important rough grass habitat (Fig. 3.4b). Overall quality of terrestrial habitat was also influential (Fig. 3.4c), where ponds had higher species richness in areas considered to be poor or moderate habitat for *T. cristatus*. Species richness was reduced as percentages of terrestrial overhang (Fig. 3.4d) and macrophyte cover increased (Fig. 3.4e), but improved with pond density (Fig. 3.4f).

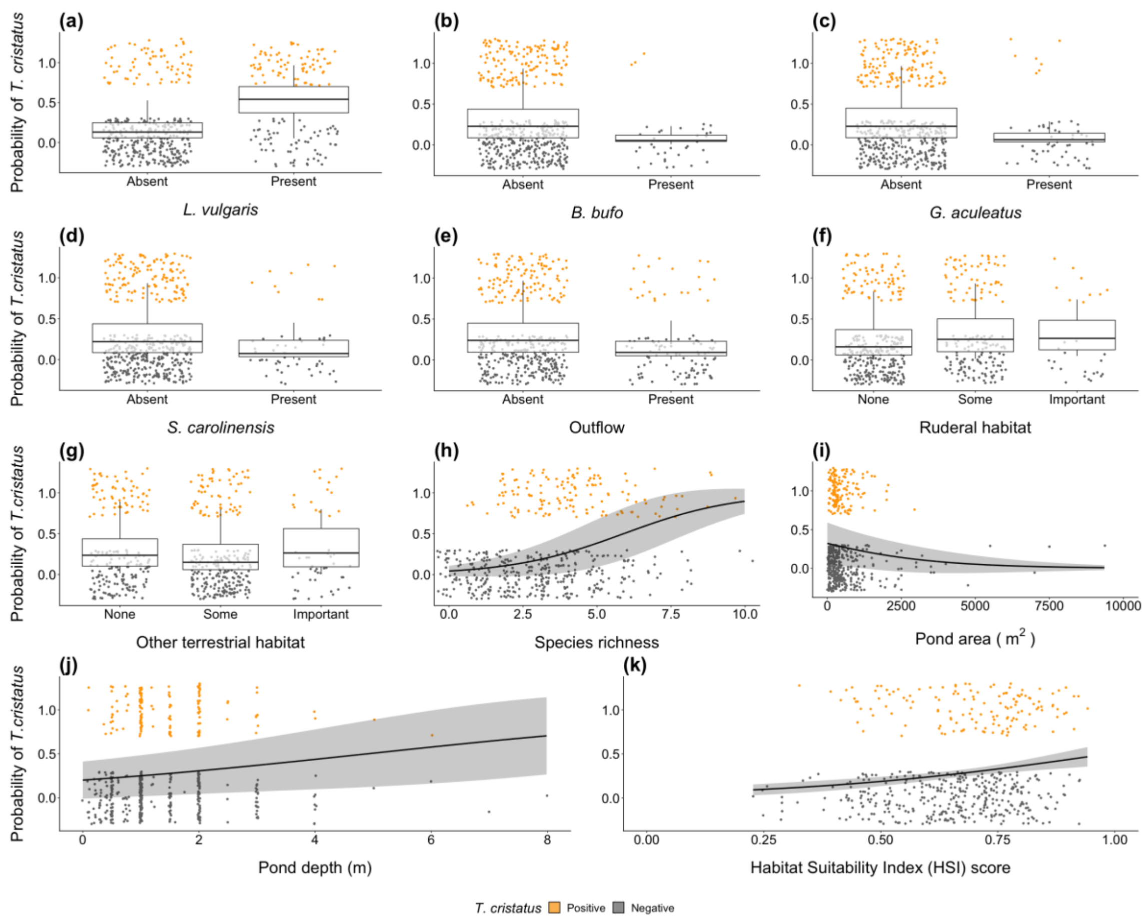


Figure 3.3: Biotic and abiotic determinants of *T. cristatus* occurrence, as predicted by the binomial GLMMs (*n* = 504 ponds): **(a)** *L. vulgaris* occurrence, **(b)** *B. bufo* occurrence, **(c)** *G. aculeatus* occurrence, **(d)** *S. carolinensis* occurrence, **(e)** pond outflow, **(f)** ruderal habitat, **(g)** other good quality terrestrial habitat, **(h)** species richness, **(i)** pond area, **(j)** pond depth, and **(k)** HSI score. The 95% CIs, as calculated using the predicted *T. cristatus* probability values and standard error for these predictions, are given for each relationship. The observed *T. cristatus* presence (orange) and absence (grey) data are also displayed as points, which have been jittered around 0 and 1 to clarify variation in point density, against the predicted relationships (boxes/lines).

Table 3.2: Summary of analyses testing for variation in vertebrate species richness across ponds (n = 504) analysed using eDNA metabarcoding, attributable to aquatic and terrestrial habitat. Results of the separate GLMM assessing variation explained by HSI score are italicised. For categorical variables with more than one level, effect size and standard error (SE) are only given for levels reported in the model summary. Test statistic is for LRT used and significant P-values (<0.05) are in bold.

Model variables	DF	Effect size (SE)	χ^2	*P*
Outflow	1		11.220	**<0.001**
Present		0.214 (0.063)		
Rough grass	2		16.715	**<0.001**
Some		-0.297 (0.074)		
None		-0.140 (0.080)		
Overall terrestrial habitat	2		8.244	**0.016**
Moderate		0.216 (0.078)		
Poor		0.115 (0.089)		
Pond density	1	0.006 (0.003)	4.564	**0.033**
Macrophyte cover	1	-0.002 (0.001)	4.117	**0.043**
Terrestrial overhang	1	-0.003 (0.001)	9.575	**0.002**
HSI score	*1*	*0.459 (0.002)*	*5.034*	***0.025***

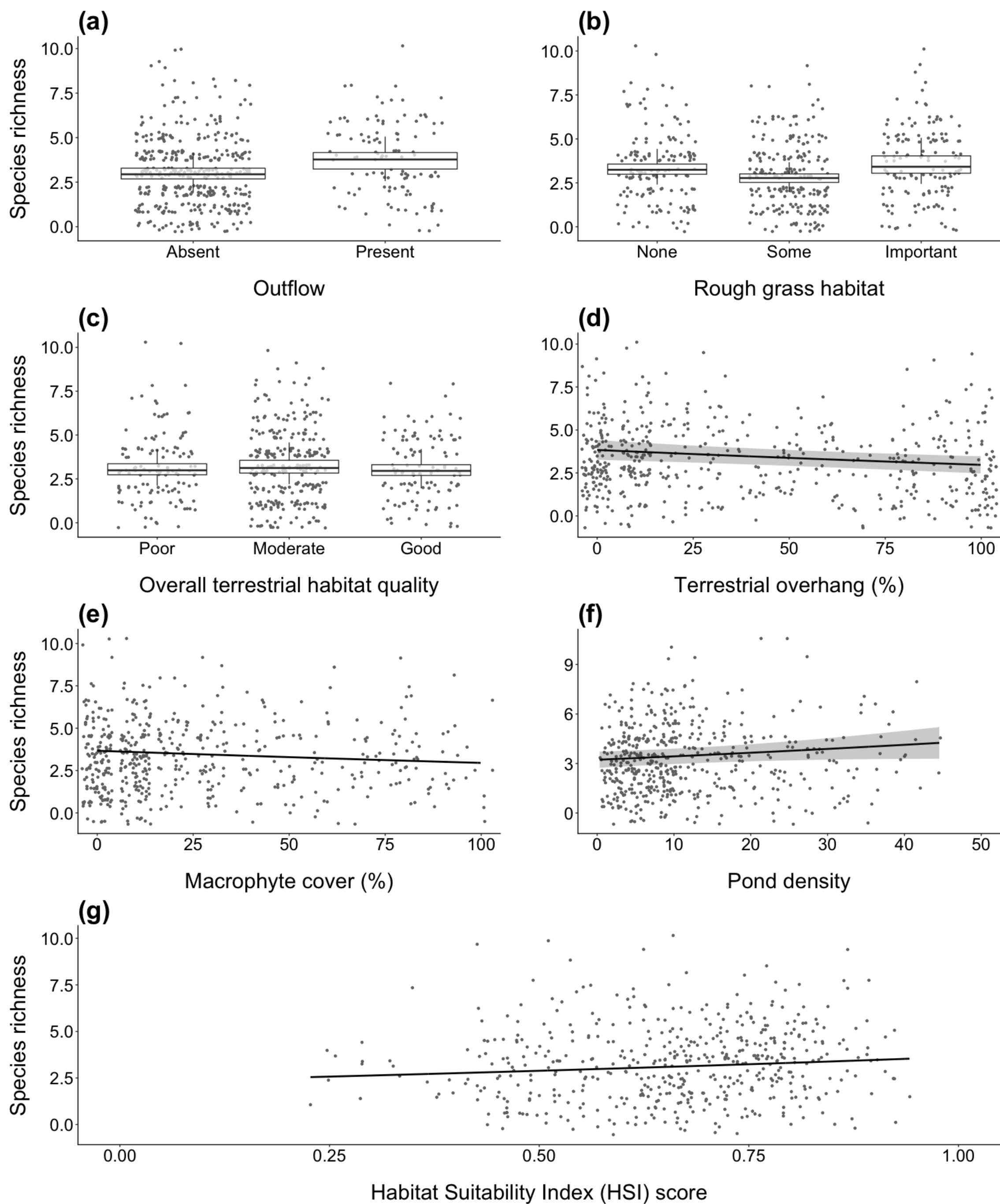


Figure 3.4: Abiotic determinants of vertebrate species richness, as predicted by the Poisson GLMM (n = 504 ponds): **(a)** outflow, **(b)** rough grass habitat, **(c)** overall quality of terrestrial habitat, **(d)** percentage of terrestrial overhang, **(e)** percentage of macrophyte cover, **(f)** pond density, and **(g)** HSI score. The 95% CIs, as calculated using the predicted species richness values and standard error for these predictions, are given for each relationship. The observed data are also displayed as points, which have been jittered around 0 and 10 to clarify variation in point density, against the predicted relationships (boxes/lines).

3.3.6 Applicability of the HSI to *T. cristatus* and community eDNA surveys

HSI score positively correlated with both *T. cristatus* occurrence (GLMM: overdispersion $\chi^2_{501} = 506.763$, $P = 0.4198$; fit $\chi^2_8 = 8.118$, $P = 0.422$, $R^2 = 4.99\%$) and vertebrate species richness (GLMM: overdispersion $\chi^2_{501} = 389.744$, $P = 0.999$; fit $\chi^2_8 = -145.12$, $P = 1.000$, $R^2 = 1.10\%$). *T. cristatus* occupancy probability (Table 3.1, Fig. 3.3k) and vertebrate species richness (Table 3.2, Fig. 3.4g) were improved in ponds with higher HSI score.

3.4 Discussion

We have ground-truthed eDNA metabarcoding as a tool for ecological hypothesis testing using the community data generated by this tool in combination with environmental metadata for ponds. We tested biotic and abiotic determinants of *T. cristatus* occupancy and wider vertebrate biodiversity as well as the appropriateness of the *T. cristatus* HSI for eDNA survey. *T. cristatus* occupancy was higher in ponds containing *L. vulgaris*, but devoid of *B. bufo*, *G. aculeatus*, and *S. carolinensis*. Ponds inhabited by *T. cristatus* were typically small, deep, absent of outflow, and surrounded by ruderal and good quality terrestrial habitat. Vertebrate species richness was higher in ponds where outflow, some rough grass habitat, and poor or moderate terrestrial habitat for *T. cristatus* were present. Species richness was lower at higher percentages of terrestrial overhang and macrophyte cover, but greater at higher pond density. The *T. cristatus* HSI was appropriate for predicting both *T. cristatus* occupancy and vertebrate species richness. Our findings demonstrate the power of eDNA metabarcoding to enhance freshwater monitoring and research by providing biodiversity data *en masse* at low cost.

3.4.1 Pondscape biodiversity

eDNA metabarcoding detected six amphibian, 14 fish, 17 bird, and 22 mammal species across 532 UK ponds. This diverse species inventory emphasises the importance of ponds as habitat for aquatic taxa, but also as stepping stones for semi-aquatic and terrestrial taxa (De Meester *et al.*, 2005; Hill *et al.*, 2018) through provision of drinking, foraging, dispersal, and reproductive opportunities (Biggs *et al.*, 2016; Klymus *et al.*, 2017b). Some species detections may be the result of eDNA transport from water bodies in the

surrounding area (Hänfling *et al.*, 2016) to ponds via inflow; however, this signifies the capacity of ponds to provide natural samples of freshwater biodiversity in the wider catchment (Deiner *et al.*, 2017; Harper *et al.*, 2019a).

3.4.2 Biotic determinants of *T. cristatus* occurrence

T. cristatus occurrence was positively associated with vertebrate species richness, which may support its status as an indicator or umbrella species for pond biodiversity (Gustafson *et al.*, 2006). *T. cristatus* were more likely to occur in ponds with higher amphibian species richness - particularly ponds containing *L. vulgaris* and absent of *B. bufo*. *T. cristatus* and *L. vulgaris* have similar habitat requirements and tend to breed in the same ponds (Skei *et al.*, 2006; Rannap *et al.*, 2009a; Denoël *et al.*, 2013), with >60% overlap reported (Rannap & Briggs, 2006). However, *L. vulgaris* can inhabit a broader range of habitat (Rannap & Briggs, 2006; Skei *et al.*, 2006) than *T. cristatus*, which depends on larger, deeper ponds with abundant macrophytes and no fish located in open, semi-rural landscapes (Denoël *et al.*, 2013). *B. bufo* inhabits fish-containing ponds (Manenti & Pennati, 2016) which may explain the negative association with *T. cristatus* as opposed to the positively associated *T. cristatus* and *L. vulgaris*. However, *T. cristatus* may also predate *B. bufo* eggs and larvae (Langton *et al.*, 2001).

T. cristatus was negatively associated with higher fish species richness, and specifically the presence of sticklebacks (*G. aculeatus* and *P. pungitius*) - fish that are common in and typical of ponds. All *T. cristatus* life stages may be predated by fishes (Langton *et al.*, 2001) and negative effects of fish presence-absence on *T. cristatus* occupancy, distribution, and abundance are repeatedly reported (Joly *et al.*, 2001; Rannap & Briggs, 2006; Skei *et al.*, 2006; Denoël & Ficetola, 2008; Rannap *et al.*, 2009a b; Hartel *et al.*, 2010; Denoël *et al.*, 2013). *G. aculeatus* predates *T. cristatus* eggs and larvae (McLee & Scaife, 1992; Jarvis, 2010), and has non-consumptive effects on *T. cristatus* embryos (Jarvis, 2010). *T. cristatus* larvae were also found to alter their behaviour when exposed to predatory *G. aculeatus* but not non-predatory *C. carassius* (Jarvis, 2012), another fish characteristic of ponds.

In our study, we detected *T. cristatus* in 50% of ponds inhabited by *C. carassius*, but <20% of ponds containing large and/or predatory fishes, e.g. *E. lucius* and *G. aculeatus*. Although fewer ponds contained *C. carassius* than *E. lucius* or *G. aculeatus*, previous research also indicates large and/or predatory fish are more detrimental to *T.*

cristatus occurrence (Skei *et al.*, 2006; Hartel *et al.*, 2010; Chan, 2011). *C. carassius* does not hinder *T. cristatus* oviposition, larval behaviour, or recruitment success (Chan, 2011; Jarvis, 2012), or pond invertebrate and macrophyte diversity (Stefanoudis *et al.*, 2017). In contrast, *C. carpio* foraging reduces invertebrate density and macrophyte cover (Maceda-Veiga, López & Green, 2017), which lowers *T. cristatus* reproductive and foraging success and heightens predator exposure (Rannap & Briggs, 2006; Gustafson *et al.*, 2006; Chan, 2011). *C. carassius* and *C. carpio* are both included among fish species assumed to negatively impact *T. cristatus* and whose presence-absence is assessed for the *T. cristatus* HSI (ARG-UK, 2010). However, it is evident that *C. carassius* does not directly predate *T. cristatus* or indirectly alter its behaviour, reproductive success, or habitat. Therefore, we advocate a systematic re-evaluation of problematic fish species for *T. cristatus* conservation.

Unexpectedly, *T. cristatus* was positively associated with waterfowl species richness, namely presence of *F. atra* and *G. chloropus*. These waterfowl species share macrophytes and macroinvertebrates as resources with amphibians, feeding on both directly (Perrow *et al.*, 1997; Paillisson & Marion, 2001; Wallau *et al.*, 2010). *F. atra* and *G. chloropus* crop emergent macrophytes to search for invertebrate prey (Paillisson & Marion, 2001; Wallau *et al.*, 2010), which may indirectly benefit *T. cristatus* foraging. Although *Fulica* spp. can also pull up submerged vegetation and damage vegetation banks (Lauridsen, Jeppesen & Andersen, 1993), diet is macrophyte-dominated in late summer and autumn (Perrow *et al.*, 1997) and unlikely to impact *T. cristatus* breeding in spring (Langton *et al.*, 2001). The positive association identified here between *T. cristatus* and these waterfowl most likely reflects a shared preference for macrophyte-rich ponds.

T. cristatus had negative associations with *P. colchicus* and *S. carolinensis*, but a positive association with *S. s. domesticus*. *T. cristatus* individuals are at risk of predation during the terrestrial as well as aquatic phases of their life history (Langton *et al.*, 2001; Gustafson *et al.*, 2011). There have been anecdotal records of pheasant predation on herpetofauna, including *T. cristatus* (Rice, 2016), which our results would support. However, the terrestrial associations identified may instead reflect land-use and indirect interaction. *T. cristatus* individuals prefer ponds surrounded by deciduous forest and pasture (Gustafson *et al.*, 2011), where *P. colchicus* and domestic animals are commonplace, over those in urban areas (Denoël & Ficetola, 2008; Hartel *et al.*, 2010), which support dense populations of *S. carolinensis*.

3.4.3 Abiotic determinants of *T. cristatus* occurrence

T. cristatus was more likely to occupy deeper ponds, but less likely to inhabit larger ponds, with outflow present, no ruderal habitat, and some other good terrestrial habitat. Consistent with our results, pond depth was previously found to positively influence *T. cristatus* occupancy (Denoël & Ficetola, 2008). Shallow ponds can be inhospitable due to drying or freezing and may contain less prey, but detrimental effects are often observed in open farmland (Denoël & Ficetola, 2008). Although our results indicate *T. cristatus* prefers larger ponds, pond area does not always influence occurrence (Maletzky, Kyek & Goldschmid, 2007; Denoël & Ficetola, 2008; Gustafson *et al.*, 2011) and was deemed a poor predictor of reproductive success (Vuorio *et al.*, 2013). *T. cristatus* has been found to utilise small and large ponds (Rannap & Briggs, 2006; Skei *et al.*, 2006); however, very small ponds (less than 124 m^2) may be unable to support all life stages, and larger ponds may contain fish and experience eutrophication from agricultural or polluted run-off (Rannap & Briggs, 2006). Effects of pond outflow (facilitated by drains, pipes or streams) are understudied, whereas inflow impacts biodiversity via polluted agricultural run-off and connections to streams and rivers containing large, predatory fish. Based on our findings that show outflow presence deters *T. cristatus* occupation, we suggest outflow minimises fluctuations in pond depth (Freshwater Habitats Trust, 2015) and alters patterns of colonisation and community structure to those of ponds which fluctuate with precipitation.

Our results support the importance of good terrestrial habitat to *T. cristatus* for shelter, foraging and dispersal (Langton *et al.*, 2001). We observed higher *T. cristatus* occurrence in ponds surrounded by ruderal or other good quality terrestrial habitat. However, the majority of terrestrial habitat variables were not retained by our model selection. Hartel *et al.* (2010) also found landscape variables, excluding urbanisation, were inadequate predictors of *T. cristatus* distribution, although their study area was mostly rural and thus optimal for amphibians. Conversely, research on more diverse landscapes found *T. cristatus* occupancy was lower in coniferous forest, yet higher in deciduous or herb-rich forest and pasture (Gustafson *et al.*, 2011; Vuorio *et al.*, 2013). In our study, the metadata available were qualitative, preventing detailed analyses on terrestrial habitat quality in relation to *T. cristatus* occupancy. Better understanding of occupancy and interactions with terrestrial species could be achieved with quantitative data on terrestrial habitat type, density, distance to ponds, and species utilisation.

Furthermore, given the metapopulation dynamics of *T. cristatus*, future research should investigate spatial drivers (e.g. land cover, pond density, climate variables, roads, rivers, elevation) of *T. cristatus* occupancy using innovative modelling approaches, e.g. individual-based models (Messager & Olden, 2018). However, acquiring this data to perform these models is a phenomenal task for large numbers of ponds across a vast landscape (Denoël & Ficetola, 2008).

3.4.4 Abiotic determinants of vertebrate species richness

Vertebrate species richness was higher in ponds where outflow was present, with poor or moderate overall terrestrial habitat for *T. cristatus*, and more ponds nearby. Conversely, species richness was lower in ponds with higher percentages of macrophyte cover and terrestrial overhang, and absent of or surrounded by some rough grass habitat. We compare our results to previous studies of aquatic species richness, but these largely focus on species assemblages or guilds, primarily macrophytes, macroinvertebrates, and amphibians. Outflow and inflow have been understudied in relation to pond biodiversity, although outflow may release harmful pollutants and pathogens (Beutel & Larson, 2015) that would accumulate in a closed pond system. Species richness increased as pond density increased, which echoes the positive relationship between pond density and macrophyte and macroinvertebrate richness observed by Gledhill, James & Davies (2008). These findings combined again highlight the importance of pond networks for aquatic and non-aquatic taxa (Hill *et al.*, 2018). Shade was identified as a principal driver of macroinvertebrate and macrophyte diversity in ponds (Sayer *et al.*, 2012). Yet, canopy and macrophyte cover positively influence amphibian species richness (Piha, Luoto & Merilä, 2007). Plentiful rough grass habitat may create more ecological niches and foraging opportunity for different vertebrates, but quantitative data on type and abundance of terrestrial habitat are needed to understand species preferences.

3.4.5 Applicability of the HSI to *T. cristatus* and community eDNA surveys

We found the HSI can predict eDNA-based *T. cristatus* occupancy at the pondscape. This contradicts previous conventional studies which deemed the index inappropriate for predicting *T. cristatus* occupancy or survival probabilities (Unglaub *et al.*, 2015). Our results also suggest some indices comprising the *T. cristatus* HSI, for example, outflow

and terrestrial habitat, represent suitability criteria for other biodiversity. An adapted HSI, designed to predict species richness, could help select areas for management and enhancement of aquatic and non-aquatic biodiversity. Nonetheless, the *T. cristatus* HSI also confers protection to pond biodiversity by identifying optimal habitat for pond creation and restoration to encourage populations of this threatened amphibian. The HSI is not without issue due to qualitative data used for score calculation and subjective estimation of indices (Oldham *et al.*, 2000). For future application of this index in *T. cristatus* eDNA survey, we recommend metabarcoding to quantify some qualitatively assessed indices (e.g. water quality via macroinvertebrate diversity, fish and waterfowl presence) alongside *T. cristatus* detection. Provided rigorous spatial and temporal sampling are undertaken, eDNA metabarcoding can also generate site occupancy data to estimate relative species abundance (Valentini *et al.*, 2016; Hänfling *et al.*, 2016).

3.4.6 Limitations of eDNA metabarcoding

eDNA metabarcoding has enormous potential to enhance freshwater research and enable ecological hypothesis testing at greater spatiotemporal scales (see section 3.4.7). However, species identifications must be scrutinised and validated against contemporary knowledge of species ecology and distribution. Customised reference sequence databases, containing only recorded or expected species at study sites, are crucial to achieve high-confidence identifications, reduce false positives, and prevent misinterpretation in metabarcoding studies (Port *et al.*, 2016; Shaw *et al.*, 2016a). Our custom reference databases were comprised of sequences obtained from the NCBI nucleotide (nt) database for vertebrate species recorded in the UK (Natural History Museum UK Species Database, 2017). However, many waterfowl species were missing or underrepresented in our UK bird database. Consequently, some taxonomic assignments obtained for waterfowl were unreliable, e.g. green-winged teal (*Anas carolinensis*). Researchers and practitioners must invest in the procurement, development, and curation of reference sequences for missing or underrepresented species to ensure quality and reliability of eDNA metabarcoding data. This is challenging due to the time and monetary investment needed to barcode DNA for the desired gene region from all taxa expected at study sites (Taberlet *et al.*, 2012; Shaw *et al.*, 2016a). Furthermore, morphotaxonomic expertise are required for curation thus molecular and conventional taxonomists must join forces to create reference sequence databases (Cowart *et al.*, 2015). Nonetheless,

voucher-linked DNA barcode reference databases for freshwater and terrestrial species are essential for accurate species identification and discovery of new species from pond eDNA samples.

3.4.7 Prospects of eDNA metabarcoding for conservation, management, and research

We have demonstrated the effectiveness of eDNA metabarcoding for landscape-scale biodiversity monitoring and ecological hypothesis testing. We investigated associations between aquatic and non-aquatic vertebrates, and combined metabarcoding with environmental metadata to revisit important ecological hypotheses at an unprecedented scale. Our findings indicate preferred habitat of a threatened amphibian. *T. cristatus* was more likely to occupy ponds where *L. vulgaris* was present, but *B. bufo*, *G. aculeatus*, and *S. carolinensis* were absent. *T. cristatus* prefers small but deep ponds that are absent of outflow, and surrounded by ruderal and good quality terrestrial habitat. These findings will guide management in the face of increasing land-use and habitat fragmentation - a poignant issue as protective legislation for *T. cristatus* in the UK is changing. Whilst conservation of threatened species and their habitat should be a priority, the bigger picture should not be ignored. eDNA metabarcoding could enhance our understanding of freshwater networks - particularly pondscapes - to enable more effective monitoring, protection, and management of aquatic and non-aquatic biodiversity. We are only now beginning to realise and explore these opportunities.

3.5 Acknowledgements

This work was funded by the University of Hull. We would like to thank Jennifer Hodgetts (Fera Science Ltd) for assisting with sample collection, and Jianlong Li (University of Hull) for primer design and advice on laboratory protocols. Tissue samples for primer validation and Sanger sequencing were provided by Andrew Buxton and Richard Griffiths (DICE, University of Kent) under licence from Natural England, and Barbara Mabel and Elizabeth Kilbride (University of Glasgow).

3.6 Data accessibility

The taxonomically assigned sequence reads used in this study were produced in Chapter 2 (see also Harper *et al*., 2018a). Archiving of sequence read data and the bioinformatics analysis is described in section 2.6. R scripts and corresponding data for this chapter are deposited in a separate GitHub repository (https://github.com/lrharper1/LRHarper_PhDThesis_Chapter3) which has been permanently archived (https://doi.org/10.5281/zenodo.2634033).

Chapter 4: Development and application of environmental DNA surveillance for the threatened crucian carp (*Carassius carassius*)

A crucian carp (*Carassius carassius*) (Linnaeus, 1758)
Photo by John Bailey

This chapter has been published as

Harper, L.R., Griffiths, N.P., Lawson Handley, L., Sayer, C.D., Read, D.S., Harper, K.J., Blackman, R.C., Li, J. & Hänfling, B. (2019) Development and application of environmental DNA surveillance for the threatened crucian carp (*Carassius carassius*). *Freshwater Biology*, **64**, 93–107.

Abstract

The crucian carp (*Carassius carassius*) is one of few fish species associated with small ponds in the UK. These populations contain genetic diversity not found in Europe and are important to conservation efforts for the species, which has declined across its range in Europe. Detection and monitoring of extant *C. carassius* populations are crucial for conservation success. Environmental DNA (eDNA) analysis could be very useful in this respect as a rapid, cost-efficient monitoring tool. We developed a species-specific quantitative PCR (qPCR) assay for eDNA surveillance of *C. carassius* to enable non-invasive, large-scale distribution monitoring. We compared fyke netting and eDNA at ponds with (N = 10) and without (N = 10) *C. carassius* for presence-absence detection. We examined biotic (*C. carassius* density represented by Catch-Per-Unit-Effort [CPUE] estimate) and abiotic influences on eDNA detection probability using a hierarchical occupancy model, and eDNA quantification using a mixed-effects model. eDNA analysis achieved 90% detection for *C. carassius* (N = 10), failing in only one pond where presence was known. CPUE estimate and conductivity had positive and negative influences on eDNA detection probability in qPCR replicates respectively. Similarly, conductivity had a negative effect on DNA copy number, whereas copy number increased with CPUE estimate. Our results demonstrate that eDNA analysis could enable detection of *C. carassius* populations in ponds and benefit ongoing conservation efforts, but imperfect species detection in relation to biotic and abiotic factors and eDNA workflow requires further investigation. Nonetheless, we have established an eDNA framework for *C. carassius* as well as sources of imperfect detection which future investigations can build upon.

4.1 Introduction

The crucian carp (*Carassius carassius*) (Fig. 4.1) is an elusive, benthic fish species popular with anglers (Copp, Warrington & Wesley, 2008b; Sayer *et al*., 2011). As one of few fish associated with small ponds, this species may have an important ecological role but its relationship with other lentic biodiversity is understudied (Copp & Sayer, 2010; Stefanoudis *et al*., 2017). Although listed as 'Least Concern' on the International Union for Conservation of Nature (IUCN) Red List of Threatened Species, the species has declined throughout its native range of Northwest and Central Europe (Copp *et al*., 2008b; Sayer *et al*., 2011), with local extinctions across the UK (Copp & Sayer, 2010). The county of Norfolk in eastern England was believed to hold abundant and widely distributed *C. carassius* populations, but research indicates heavy (~75%) declines in this region (Sayer *et al*., 2011). Declines of *C. carassius* throughout its range are due to habitat loss (Copp *et al*., 2008b; Sayer *et al*., 2011), species displacement by the invasive gibel carp (*Carassius gibelio*, Copp *et al*., 2008b; Tarkan *et al*., 2009; Sayer *et al*., 2011), and genetic introgression through hybridisation (Hänfling *et al*., 2005). Indeed, Sayer *et al*. (2011) observed only 50% of *C. carassius* ponds to be uninhabited by goldfish (*Carassius auratus*), common carp (*Cyprinus carpio*), or their hybrids with *C. carassius*.

Prior to the 1970's, *C. carassius* were thought to have been introduced to the UK alongside *C. carpio* and were classed as non-native (Maitland, 1972). Wheeler (1977) deemed the species native to southeast England based on archaeological evidence and a historic distribution that mirrored native cyprinids. Conservation organisations (e.g. English Nature, Environment Agency) later recognised *C. carassius* as native and threatened (Smith & Moss, 1994; Environment Agency, 2003), but recent genetic evidence supports anthropogenic introduction of *C. carassius* to the UK during the 15th century (Jeffries *et al*., 2017). Nonetheless, many introduced species in the UK are now naturalised, and several provide ecological and economical benefits (Manchester & Bullock, 2000). Evidence suggests that *C. carassius* is characteristic of small, plant-dominated, high-quality ponds (Copp *et al*., 2008b; Sayer *et al*., 2011; Stefanoudis *et al*., 2017), and English populations contain a substantial proportion of the overall genetic diversity for the species across Europe. English *C. carassius* populations may buffer species displacement by *C. gibelio* at the European level (Jeffries *et al*., 2017), but are threatened by hybridisation with *C. auratus* and possible displacement (Hänfling *et al*., 2005; Tarkan *et al*., 2009) as well as anthropogenic activity (Copp, Černý & Kováč,

2008a).

Figure 4.1: A crucian carp (*Carassius carassius*) individual (a) and examples of the study ponds (b-d). Photograph **(a)** by John Bailey, photographs **(b)** and **(c)** by Carl Sayer, and photo **(d)** by Sacha Dench.

In 2010, *C. carassius* was designated as a Biodiversity Action Plan (BAP) species in the county of Norfolk (Copp & Sayer, 2010; Sayer *et al*., 2011). To meet the BAP aims, local conservation efforts have included species reintroduction, pond restoration, and eradication of *C. auratus* (Sayer *et al*., 2011). However, current distribution records are unreliable as individuals are frequently misidentified as the feral brown variety of *C. auratus* due to high physical similarity (Copp *et al*., 2008a; Tarkan *et al*., 2009), and many pond populations are mixtures of true *C. carassius* and *C. carassius* x *C. auratus* hybrids (Hänfling *et al*., 2005). Consequently, distribution maps have been called into question and further monitoring is needed to ensure long-term success of established and reintroduced *C. carassius* populations (Copp *et al*., 2008a; Tarkan *et al*., 2009).

Primarily, *C. carassius* are surveyed using fyke netting or electrofishing, but these

methods can be costly and time-consuming. Environmental DNA (eDNA) analysis offers a potentially rapid and cost-effective approach to fish monitoring (Jerde *et al*., 2011; Sigsgaard *et al*., 2015; Hänfling *et al*., 2016; Wilcox *et al*., 2016; Hinlo *et al*., 2017a). Species are identified using DNA deposited in the environment by individuals via secretions, excretions, gametes, blood, or decomposition (Lawson Handley, 2015). eDNA has been applied worldwide to survey for invasive freshwater fish (Jerde *et al*., 2011; Keskin, 2014; Robson *et al*., 2016; Hinlo *et al*., 2017a), and is now used routinely to monitor Asian carp (*Hypophthalmichthys* spp.) invasion in the Great Lakes, USA (Farrington *et al*., 2015). A quantitative PCR (qPCR) assay targeting *C. carassius* was also published in the context of early warning invasion monitoring for fish species that may arrive in Canada (Roy *et al*., 2017), but was only tested on tissue-derived DNA. Of equal importance to invasion monitoring, eDNA analysis has enhanced surveys for threatened and endangered freshwater fish (Sigsgaard *et al*., 2015; Schmelzle & Kinziger, 2016; Piggott, 2016; Bylemans *et al*., 2017).

eDNA analysis has been conducted with conventional PCR (PCR) (Ficetola *et al*., 2008; Jerde *et al*., 2011), but qPCR and droplet digital PCR (ddPCR) are suggested to perform better, suffer less from inhibition, and enable abundance or biomass estimation (Nathan *et al*., 2014). However, these estimates can be inconsistent across habitats and target organisms. In flowing water, Hinlo *et al*. (2017a) found no relationship between DNA copy number and conventional density estimates of *C. carpio*, yet Takahara *et al*. (2012) observed a positive association between *C. carpio* biomass and eDNA concentration in ponds. Environmental variables play a substantial role in abundance/biomass estimation by influencing the ecology of eDNA (Barnes *et al*., 2014). Variables examined have included temperature, pH, salinity, conductivity, anoxia, sediment type, and ultraviolet (UV) light (Takahara *et al*., 2012; Barnes *et al*., 2014; Keskin, 2014; Pilliod *et al*., 2014; Strickler *et al*., 2015; Robson *et al*., 2016; Buxton *et al*., 2017a, b; Stoeckle *et al*., 2017; Weltz *et al*., 2017; Goldberg *et al*., 2018). However, these variables are not always measured and only a handful of studies have assessed their effects in ponds (Takahara *et al*., 2012; Buxton *et al*., 2017a, b; Goldberg *et al*., 2018).

In this study, we developed a species-specific qPCR assay for the threatened *C. carassius*. We evaluated presence-absence detection with eDNA compared to fyke netting, and investigated the influence of biotic and abiotic factors on eDNA detection and quantification. We hypothesised that: (1) eDNA and fyke netting would provide comparable presence-absence records for *C. carassius*, and (2) eDNA detection and

quantification would be influenced by *C. carassius* density, temperature, pH, conductivity, surface dissolved oxygen, macrophyte cover, and tree shading. We provide an eDNA framework for *C. carassius* monitoring which holds promise for routine survey.

4.2 Materials and methods

4.2.1 Study sites

We studied 10 ponds with confirmed *C. carassius* presence at different densities and 10 fishless ponds in Norfolk (Fig. 4.2). This region is low-lying (<100 m above sea level) and mainly agricultural. All study ponds were selected to be small (<40 m in max. dimension), shallow (<2 m), macrophyte-dominated, with a largely open-canopy and thus minimal shading of the water surface. Ponds were largely surrounded by arable fields, excluding one located in woodland. No specific permits were required for sampling but relevant landowner permissions were obtained.

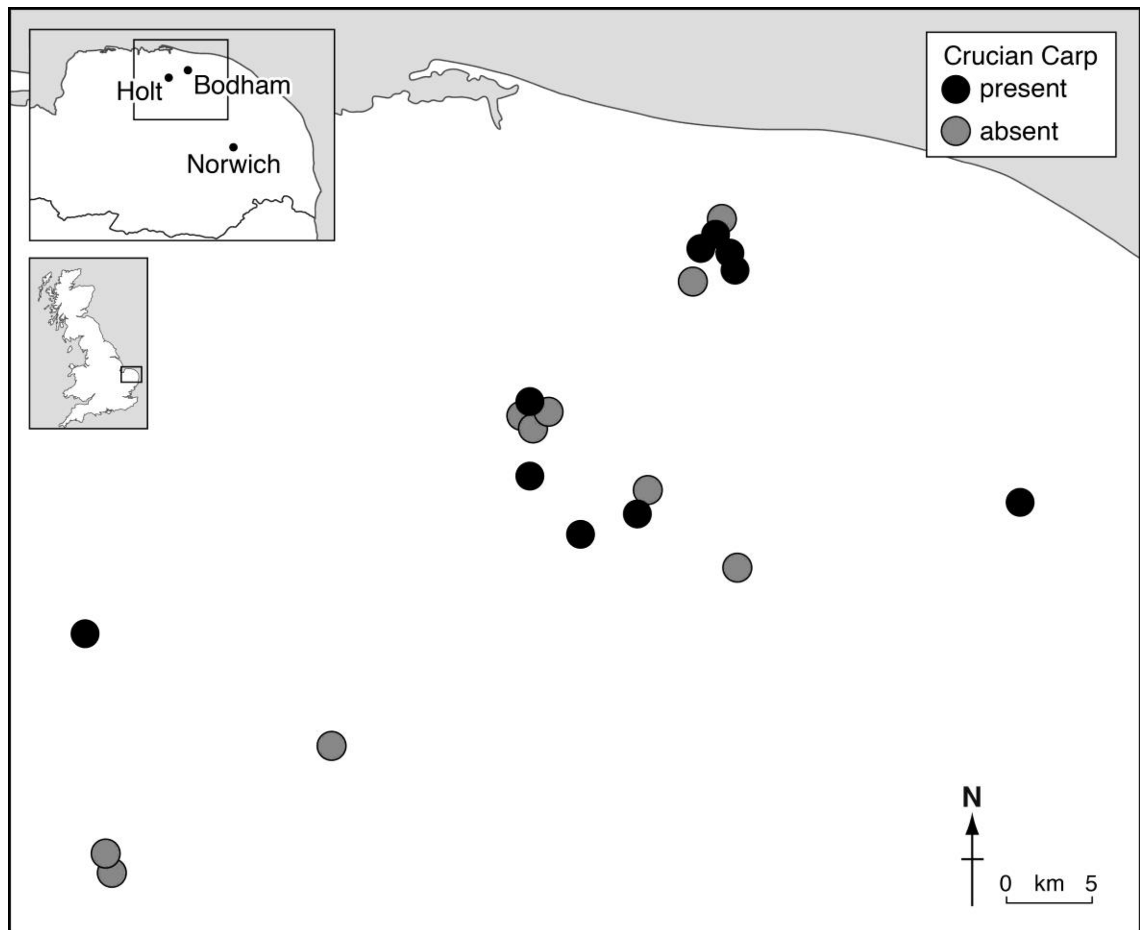


Figure 4.2: Map of pond locations in North Norfolk, eastern England, showing the distribution of ponds containing crucian carp (*C. carassius*, black dots) and ponds where the species is absent (grey dots).

4.2.2. Conventional survey

C. carassius presence-absence was confirmed at each pond by fyke netting between 2010 and 2016. Bar two ponds surveyed in 2013 and 2015, all *C. carassius* ponds were last surveyed in 2016. Where possible, double-ended fyke nets were set perpendicular to the bank or to beds of aquatic vegetation and exposed overnight (for c. 16 h), with the number of fyke nets set being proportional to pond size. This provided Catch-Per-Unit-Effort (CPUE) estimates of relative densities, which are the number of fish captured per fyke net per 16 h exposure. Environmental data were collected between May and August from 2010 to 2017. Conductivity, pH, surface dissolved oxygen, and water temperature were measured with a HACH HQ30d meter (Hach Company, CO, USA), and alkalinity was

determined by sulphuric-acid titration using a HACH AL-DT kit (Hach Company, CO, USA). Percentages of macrophyte cover and shading of ponds by trees and scrub were estimated visually.

4.2.3 eDNA sampling, capture and extraction

Five 2 L surface water samples were collected from the shoreline of each pond using sterile Gosselin™ HDPE plastic bottles (Fisher Scientific UK Ltd, UK) and disposable gloves. Samples were taken at equidistant points around the pond perimeter where access permitted. All ponds without *C. carassius* were sampled on 22nd August 2016. Water samples were transported on ice in sterile coolboxes to the Centre for Ecology & Hydrology, Wallingford, stored at 4 °C, and vacuum-filtered within 24 hours of collection. Coolboxes were sterilised using 10% v/v chlorine-based commercial bleach (Elliott Hygiene Ltd, UK) solution and 70% v/v ethanol solution before ponds containing *C. carassius* were sampled on 25th August 2016. Samples were handled in the same way as those from fishless ponds. For each pond, a full process blank (1 L molecular grade water) was taken into the field and stored in coolboxes with samples. Blanks were filtered and extracted alongside pond samples to identify contamination.

Where possible, the full 2 L of each sample was vacuum-filtered through sterile 0.45 μm cellulose nitrate membrane filters with pads (47 mm diameter; Whatman, GE Healthcare, UK) using Nalgene™ filtration units. One hour was allowed for each sample to filter but if filters clogged during this time, a second filter was used. After 2 L had been filtered or one hour had passed, filters were removed from pads using sterile tweezers, placed in sterile 47 mm petri dishes (Fisher Scientific UK Ltd, UK), sealed with parafilm (Sigma-Aldrich®, UK), and stored at -20 °C. The total volume of water filtered and the number of filters used per sample were recorded for downstream analysis (Table S4.1). After each round of filtration (samples and blanks from two ponds), all equipment was sterilised in 10% v/v chlorine-based commercial bleach (Elliott Hygiene Ltd, UK) solution for 10 minutes, immersed in 5% v/v MicroSol detergent (Anachem, UK), and rinsed with purified water.

All filters were transported on ice in a sterile coolbox to the University of Hull and stored at -20 °C until DNA extraction one week later. DNA was isolated from filters using the PowerWater® DNA Isolation Kit (MO BIO Laboratories, CA, USA) following the manufacturer's protocol in a dedicated eDNA facility at the University of Hull,

devoted to pre-PCR processes with separate rooms for filtration, DNA extraction, and PCR preparation of environmental samples. Duplicate filters from the same sample were co-extracted by placing both filters in a single tube for bead milling. Eluted DNA (100 μL) concentration was quantified on a Qubit™ 3.0 fluorometer using a Qubit™ dsDNA HS Assay Kit (Invitrogen, UK). DNA extracts were stored at -20 °C until further analysis.

4.2.4 Assay design, specificity and sensitivity

We designed a novel qPCR assay to target a 118 bp amplicon (73 bp excluding primers) within the mitochondrial cytochrome *b* (*cytb*) gene, specific to *C. carassius*. *C. carassius* sequences from Jeffries *et al*. (2016) were aligned using MAFFT in AliView (Larsson, 2014) to sequences downloaded from the NCBI nucleotide (nt) database for 23 closely related species of European freshwater fish (Table S4.2), and a consensus sequence for each species was identified (Fig. 4.3). Sequences were visually compared to maximise nucleotide mismatches between *C. carassius* and non-target species, particularly *C. auratus* and *C. carpio*, and minimise theoretical risk of non-specific amplification. Mismatches in primer regions were maximised over the probe region to increase specificity (Wilcox *et al*., 2013). Species-specific primers CruCarp_CytB_984F (5'-AGTTGCAGATATGGCTATCTTAA-3') and CruCarp_CytB_1101R (5'-TGGAAAGAGGACAAGGAATAAT-3'), and corresponding probe CruCarp_CytB_1008Probe (FAM 5'-ATGGATTGGAGGCATACCAGTAGAACACC-3' BHQ1) were selected on this basis.

Primers without probe were tested *in silico* using ecoPCR (Ficetola *et al*., 2010) against a custom, phylogenetically curated reference database that was constructed for eDNA metabarcoding of lake fish communities in Windermere, Lake District National Park, England, which contains 67 freshwater fish species confirmed or potentially present in the UK (Hänfling *et al*., 2016). Parameters set allowed a 50-150 bp fragment and maximum of three mismatches between each primer and each sequence in the reference database. Specificity of primers (without probe) was also tested against the full NCBI nucleotide (nt) database using Primer-BLAST (Ye *et al*., 2012) with default settings.

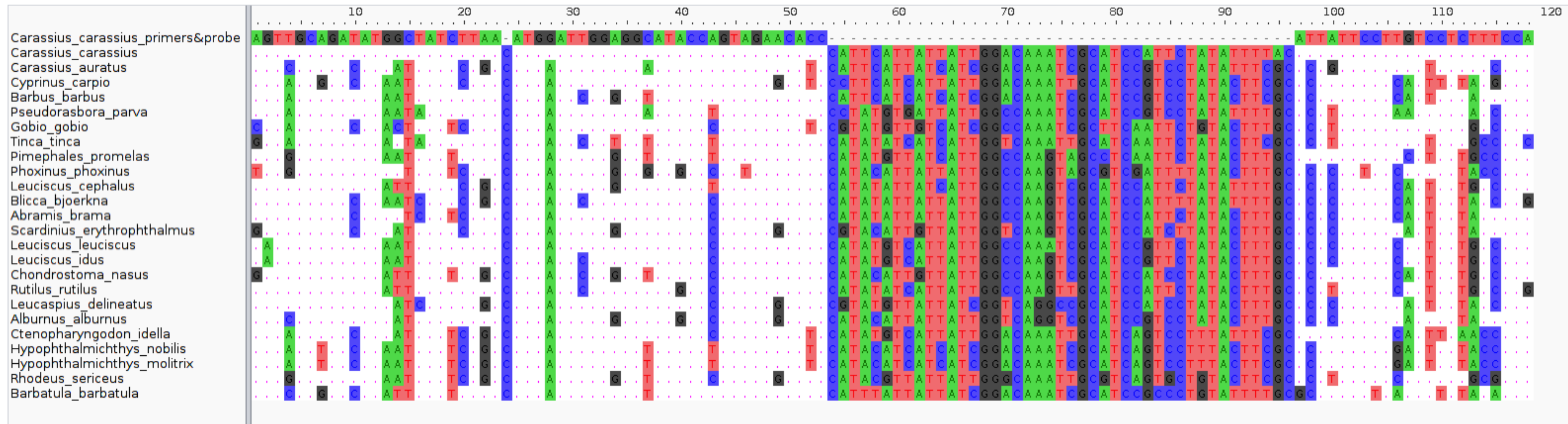


Figure 4.3: Alignment of consensus sequences for a region of the mitochondrial cytochrome *b* (*cytb*) gene in 24 European freshwater fishes, including *C. carassius*. Species-specific primers and probe for *C. carassius* are given on the first line. Consensus with primer and probe sequence across species is highlighted in white whereas mismatches are coloured by nucleotide base.

The primers were tested with PCR, following which primer and probe concentrations, standard curve preparation, and cycling conditions for qPCR were optimised (Appendix 4). All subsequent qPCR analyses were performed using the conditions detailed in section 4.2.5. Primers and probe were validated *in vitro* using tissue DNA (standardised to 1 ng/µL) from fin clips of 10 non-target species (1 UK individual per species) related to *C. carassius* (Table S4.3, Figures S4.1-4.3). The positive control and No Template Control (NTC) were *C. carassius* DNA and molecular grade water (Fisher Scientific UK Ltd, UK) respectively. The limits of detection (LOD, the lowest concentration where at least one technical replicate amplified *C. carassius* DNA) and quantification (LOQ, the concentration at which all technical replicates consistently amplified *C. carassius* DNA) (Agersnap *et al*., 2017) were established using the qPCR standards (10^6 to 1 copy/µL, Figure S4.4). Five technical replicates were performed for standards, controls, and samples in tests of assay specificity and sensitivity.

4.2.5 Detection and quantification of *C. carassius* eDNA

All qPCR reactions were prepared in a UV and bleach (Elliott Hygiene Ltd, UK) sterilised laminar flow hood in the dedicated eDNA facility at the University of Hull. Reactions were performed in a total volume of 20 µL, consisting of 2 µL of template DNA, 1 µL of each primer (Forward 900 nM, Reverse 600 nM), 1 µL of probe (125 nM) (Integrated DNA Technologies, Belgium), 10 µL of TaqMan® Environmental Master Mix 2.0 (Life Technologies, CA, USA), and 5 µL molecular grade water (Fisher Scientific UK Ltd, UK). Once eDNA samples and three NTCs were added to each 96-well plate, the plate was sealed and transported to a separate laboratory on a different floor for addition of the standard curve and three positive controls (*C. carassius* DNA, 0.01 ng/µL) in a UV and bleach sterilised laminar flow hood.

Our standard curve was a synthesised 500 bp gBlocks® Gene Fragment (Integrated DNA Technologies, Belgium) based on GenBank accessions (KT630374 - KT630380) for *C. carassius* from Norfolk (Jeffries *et al*., 2016). Copy number for the gBlocks® fragment was estimated by multiplying Avogadro's number by the number of moles. We performed a 10-fold serial dilution of the gBlocks® fragment to generate a 6-point standard curve that ranged from 10^6 to 10 copies/µL. eDNA samples were compared to these known concentrations for quantification (Hinlo *et al*., 2017a). Each standard was

replicated five times on each qPCR plate. Similarly, five technical replicates were performed for every sample and full process blank from each pond.

After addition of standards and positive controls, plates were again sealed and transported to a separate laboratory on a different floor where qPCR was conducted on a StepOnePlus™ Real-Time PCR system (Life Technologies, CA, USA). Thermocycling conditions consisted of incubation for 5 min at 50 °C, a 10 min denaturation step at 95 °C, followed by 60 cycles of denaturation at 95 °C for 15 s and annealing at 60 °C for 1 min. We used 60 cycles for consistency with optimisation tests, but cycling could be reduced to 45 cycles for subsequent applications (see Appendix 4.1). A small-scale comparison of eDNA detection and concentration using PCR and qPCR was also conducted (Appendix 4.1).

Amplifications were considered positive detections if the exponential phase occurred within 45 reaction cycles as the mean C_q value was 40.07 for the LOD (1 copy/µL). A pond was considered positive for *C. carassius* if two or more of the five technical replicates from a sample returned positive, or more than one sample returned any positive technical replicates (Goldberg *et al*., 2016). False negatives were obtained for one pond, therefore all samples were tested for inhibition by spiking duplicate qPCR reactions with a known concentration (1000 copies/µL) of synthetic *C. carassius* template (Jane *et al*., 2015).

4.2.6 DNA sequencing

Non-target DNA extracts and full-process blanks that amplified with qPCR were Sanger sequenced alongside a representative eDNA sample from each positive pond (n = 9) to confirm sequence identity. Purification and sequencing was performed by Macrogen Europe (Amsterdam, The Netherlands) in triplicate in the forward direction. Sequences were edited using CodonCode Aligner (CodonCode Corporation, MA, USA) with default settings. Sequences were then manually aligned in AliView (Larsson, 2014) and poor quality sequences were discarded (Figure S4.5). Primers were removed from remaining sequences, and sequences identified against the full NCBI nucleotide (nt) database using the NCBI BLASTn tool.

4.2.7 Data analysis

Technical replicates for each qPCR standard that differed by >0.5 C_q from the average of the five technical replicates performed were discarded to minimise bias induced by pipetting error. All technical replicates for eDNA samples were retained, and those which failed to amplify were classed as 0 copies/µL (Goldberg *et al.*, 2016). The C_q values for each set of technical replicates were averaged and quantified to provide a single DNA copy number for each sample. Samples with no positive amplifications were assigned a DNA copy number of zero. DNA copy numbers of samples were then averaged to generate a single DNA copy number for each pond.

All subsequent data analyses were performed in the statistical programming environment R v.3.4.2 (R Core Team, 2017). Effects of water volume filtered, number of filters used, and water sample content on DNA copy number of samples were tested and reported in Appendix 4.2 (Figs. S4.6, S4.7). Results and discussion of the PCR-qPCR comparison are also reported in Appendix 4.2 (Table S4.4; Fig. S4.8). The R package eDNAoccupancy v0.2.0 (Dorazio & Erickson, 2017) was used to fit a Bayesian, multi-scale occupancy model to estimate eDNA detection probability at sites where *C. carassius* was confirmed as present by fyke netting. Existing eDNA literature was used to identify biotic and abiotic factors reported to affect eDNA detection, persistence and degradation, and construct hypotheses regarding their effects on eDNA detection probability in water samples (θ), and eDNA detection probability in qPCR replicates (p). No covariates were included at the site level (ψ) as ponds were occupied by *C. carassius* and eDNA should have been present. At the sample level, more individuals (reflected by CPUE) should increase eDNA concentration and improve detection. Temperature can increase physical, metabolic, or behavioural activity of organisms resulting in more eDNA release, breakdown, and degradation (Takahara *et al.*, 2012; Pilliod *et al.*, 2014; Strickler *et al.*, 2015; Lacoursière-Roussel, Rosabal & Bernatchez, 2016b; Robson *et al.*, 2016; Bylemans *et al.*, 2017; Buxton *et al.*, 2017b). Links established between eDNA and pH support greater detectability, concentration, and persistence of eDNA in more alkaline waters (Barnes *et al.*, 2014; Strickler *et al.*, 2015; Goldberg *et al.*, 2018). Conductivity relates to Total Dissolved Solids (TDS) and sediment type, which can impair eDNA detection due to release of inhibitory substances and their capacity to bind DNA (Buxton *et al.*, 2017a; Stoeckle *et al.*, 2017). Vegetated ponds reduce UV exposure thereby preserving eDNA (Barnes *et al.*, 2014), and are susceptible to terrestrialisation which can

create anoxic conditions that may slow eDNA degradation (Barnes *et al*., 2014; Pilliod *et al*., 2014; Weltz *et al*., 2017). At the qPCR replicate level, covariates again included CPUE as higher eDNA concentration should improve amplification success and consistency, whereas conductivity may indicate inhibitory substances that cause amplification failure.

Prior to modelling, all environmental variables were assessed for collinearity using Spearman's correlation coefficient and Variance Inflation Factors (VIFs) calculated using the R package car v2.1-6 (Fox & Weisberg, 2011). Variables were considered collinear and removed if r >0.3 and VIF >3 (Zuur *et al*., 2009), following which candidate variables (CPUE, conductivity, pH, and percentage of macrophyte cover) were centred and scaled to have a mean of 0 and standard deviation of 1. We constructed 64 models which assumed a constant probability of eDNA occurrence at the site level, and different covariate combinations at the sample and qPCR replicate levels. Models were ranked (Table S4.5) according to posterior predictive loss criterion (PPLC) under squared-error loss and the widely applicable information criterion (WAIC). The model with the best support was selected for comparison to the null model without covariates at the entire sampling hierarchy.

We examined the influence of biotic and abiotic factors on eDNA quantification using a Generalized Linear Mixed Model (GLMM) within the R package glmmTMB v0.2.0 (Brooks *et al*., 2017). Collinearity was assessed as above, leaving CPUE, pH, conductivity, and percentage of macrophyte cover as explanatory variables. Pond was modelled as a random effect to account for spatial autocorrelation in our data set and the influence of other properties inherent to each pond, whereas all other explanatory variables were fixed effects. A Poisson distribution was specified as the nature of the response variable (DNA copy number) was integer count data. Validation checks were performed to ensure all model assumptions were met and absence of overdispersion (Zuur *et al*., 2009). Model fit was assessed visually and with the Hosmer and Lemeshow Goodness of Fit Test (Hosmer & Lemeshow, 2000) using the R package ResourceSelection v0.3-0 (Lele *et al*., 2014). Model predictions were obtained using the *predict* function and upper and lower 95% CIs were calculated from the standard error of the predictions. All values were bound in a new data frame and model results plotted for evaluation using the R package ggplot2 v2.2.1 (Wickham, 2016).

4.3 Results

4.3.1 Assay specificity and sensitivity

Only *C. carassius* amplified in ecoPCR, confirming primer specificity. Non-target species returned by primer-BLAST against the full NCBI nucleotide (nt) database were *Barilius bakeri* (a Cyprinid fish restricted to India, 6 mismatches), *Naumovozyma dairensis* (fungi, 8 mismatches), and *Medicago trunculata* (plant, 8 mismatches). Our probe sequence could not be included *in silico* but would likely increase specificity. Tissue extracts from common rudd (*Scardinius erythrophthalmus*) and European chub (*Squalius cephalus*) included in qPCR assay specificity tests were amplified by primers and probe, but possessed low DNA copy number (<10 copies/µL). In a later test, *C. carpio* DNA also amplified (<10 copies/µL). However, no amplification was observed for NTCs, fresh tissue extracts obtained from *S. erythrophthalmus* and *S. cephalus*, or eDNA samples from locations where *C. carassius* were absent and these species were present (data not shown). DNA sequencing confirmed cross-contamination of reference material, where sequences were either identified as *C. carassius* or of poor quality (Table 4.1). Our assay was highly sensitive with a LOD of 1 copy/µL and LOQ of 10 copies/µL.

Table 4.1: Top NCBI BLASTn hit for Sanger sequences obtained from target DNA (tissue extracts and synthetic gBlocks® Gene Fragment), non-target tissue DNA extracts, full process blanks, and representative eDNA samples that amplified during qPCR. Sample descriptions marked with '!' indicate a poor quality, discarded sequence.

Sample	Description	Query length	Coverage	E value	Identity	Accession
CrucianCarp-01	*Carassius carassius*	37	100%	3E-09	100%	KR131843.1
CrucianCarp-02	*Carassius carassius*	37	100%	3E-09	100%	KR131843.1
CrucianCarp-03	!					
Gblock-100copies-01	*Carassius carassius*	37	100%	3E-09	100%	KR131843.1
Gblock-100copies-02	!					
Gblock-100copies-03	!					
Rudd-JL-01	*Carassius carassius*	38	100%	9E-10	100%	KR131843.1
Rudd-JL-02	*Carassius carassius*	38	100%	9E-10	100%	KR131843.1
Rudd-JL-03	!					
Rudd-PS-01	!					
Rudd-PS-02	!					
Rudd-PS-03	!					
Chub-PS-01	!					
Chub-PS-02	!					
Chub-PS-03	!					
Chub-JL-01	!					
Chub-JL-02	!					
Chub-JL-03	!					

CommonCarp-01	!					
CommonCarp-02	!					
CommonCarp-03	!					
POFA4-B-01	!					
POFA4-B-02	!					
POFA4-B-03	!					
GUES1-5-01	*Carassius carassius*	37	100%	3E-09	100%	KR131843.1
GUES1-5-02	*Carassius carassius*	41	100%	1E-07	95%	KR131843.1
GUES1-5-03	*Carassius carassius*	41	100%	2E-11	100%	KR131843.1
MYST-3-01	*Carassius carassius*	46	100%	4E-14	100%	KR131843.1
MYST-3-02	!					
MYST-3-03	*Carassius carassius*	41	100%	1E-07	95%	KR131843.1
SKEY1-4-01	*Carassius carassius*	35	100%	4E-08	100%	KR131843.1
SKEY1-4-02	!					
SKEY1-4-03	*Carassius carassius*	37	100%	3E-09	100%	KR131843.1
OTOM-4-01	!					
OTOM-4-02	!					
OTOM-4-03	*Carassius carassius*	37	100%	3E-09	100%	KR131843.1
POHI-2-01	*Carassius carassius*	41	100%	2E-11	100%	KR131843.1
POHI-2-02	!					
POHI-2-03	*Carassius carassius*	37	100%	3E-09	100%	KR131843.1
RAIL-4-01	!					
RAIL-4-02	*Carassius carassius*	38	100%	9E-10	100%	KR131843.1
RAIL-4-03	*Carassius carassius*	46	100%	4E-14	100%	KR131843.1
WADD3-4-01	*Carassius carassius*	25	96%	0.034	100%	KR131843.1

WADD3-4-02	!					
WADD3-4-03	*Carassius carassius*	38	100%	9E-10	100%	KR131843.1
CAKE-1-01	!					
CAKE-1-02	!					
CAKE-1-03	!					
POFA4-5-01	!					
POFA4-5-01	!					
POFA4-5-01	!					

4.3.2 qPCR analysis

The qPCR assay had an average amplification efficiency of 93.61% (range 79.61-102.49%) and an average R^2 value of 0.998 (range 0.995-0.999) for the standard curve. Only one plate had a qPCR efficiency lower than 90% but the standard curve quantified as expected, thus qPCR was not repeated. No amplification occurred in NTCs, but the full process blank for one pond (POFA4) amplified (<10 copies/µL). This was the only contaminated blank as the blank for pond POHI filtered alongside POFA4 and POHI samples, and blanks downstream of these samples did not amplify. Partial inhibition (<1000 copies/µL) occurred in a single sample from four different ponds: PYES2 (*C. carassius* absent), RAIL, POHI, and GUES1 (*C. carassius* present). However, partially inhibited samples all possessed >0 copies/µL when originally tested, and copy number did not differ substantially (higher in one instance) from other samples belonging to the same pond (Table S4.1). Consequently, partial inhibition did not influence detectability in our study, and problematic samples were not treated for inhibition and qPCRs were not repeated.

4.3.3 Presence-absence detection

eDNA surveillance detected *C. carassius* in 90% of the study ponds with confirmed presence (n = 10). Sanger sequencing of representative samples confirmed species

identity as *C. carassius* (Table 4.1). eDNA failed entirely in one pond (CHIP) that contained a sizeable *C. carassius* population (CPUE = 60.50), but samples from CHIP were not inhibited. *C. carassius* DNA was not detected at any sites where the species was absent.

4.3.4 Factors influencing eDNA detection and quantification

The occupancy model with the best support included CPUE and conductivity as covariates of eDNA detection probability in qPCR replicates (p). The model did not include any covariates of eDNA occurrence probability at sites (ψ) or eDNA detection probability in water samples (θ). Estimates of eDNA detection probability in a qPCR replicate ranged between 0.12 to 1.00 (Table 4.2), where *C. carassius* CPUE (parameter estimate = 1.357) and conductivity (parameter estimate = -2.112) played positive and negative roles in eDNA availability respectively (Figures 4.4a, b). The GLMM identified CPUE (0.020 ± 0.007, χ^2_1 = 5.426, P = 0.020) and conductivity (-0.007 ± 0.002, χ^2_1 = 8.709, P = 0.003) as significant predictors of DNA copy number, where DNA copy number was greater at higher CPUE (Figure 4.5a) but decreased as conductivity increased (Figure 4.5b).

Table 4.2: Bayesian estimates of *C. carassius* eDNA occurrence probability at a pond (ψ), eDNA detection probability in a water sample (θ), and eDNA detection probability in a qPCR replicate (p). Posterior median and 95% credible interval (CI) are given for each parameter of the occupancy model. The corresponding catch-per-unit-effort estimate (CPUE) is given for each pond.

Pond	*C. carassius* (Y/N)	CPUE estimate	ψ		θ		p	
			Posterior median	95% CI	Posterior median	95% CI	Posterior median	95% CI
CAKE	Y	43.00	0.90	0.62 - 1.00	0.83	0.70 - 0.92	0.14	0.05 - 0.33
CHIP	Y	60.50	0.90	0.62 - 1.00	0.83	0.70 - 0.92	0.12	0.03 - 0.36
GUES1	Y	121.75	0.90	0.62 - 1.00	0.83	0.70 - 0.92	0.98	0.86 - 1.00
MYST	Y	6.17	0.90	0.62 - 1.00	0.83	0.70 - 0.92	0.93	0.86 - 0.98
OTOM	Y	14.67	0.90	0.62 - 1.00	0.83	0.70 - 0.92	0.96	0.91 - 0.99
POFA4	Y	13.67	0.90	0.62 - 1.00	0.83	0.70 - 0.92	0.89	0.81 - 0.95
POHI	Y	7.25	0.90	0.62 - 1.00	0.83	0.70 - 0.92	0.44	0.28 - 0.59

RAIL	Y	58.17	0.90	0.62 - 1.00	0.83	0.70 - 0.92	1.00	0.99 - 1.00
SKEY1	Y	31.38	0.90	0.62 - 1.00	0.83	0.70 - 0.92	1.00	1.00 - 1.00
WADD3	Y	126.00	0.90	0.62 - 1.00	0.83	0.70 - 0.92	1.00	1.00 - 1.00

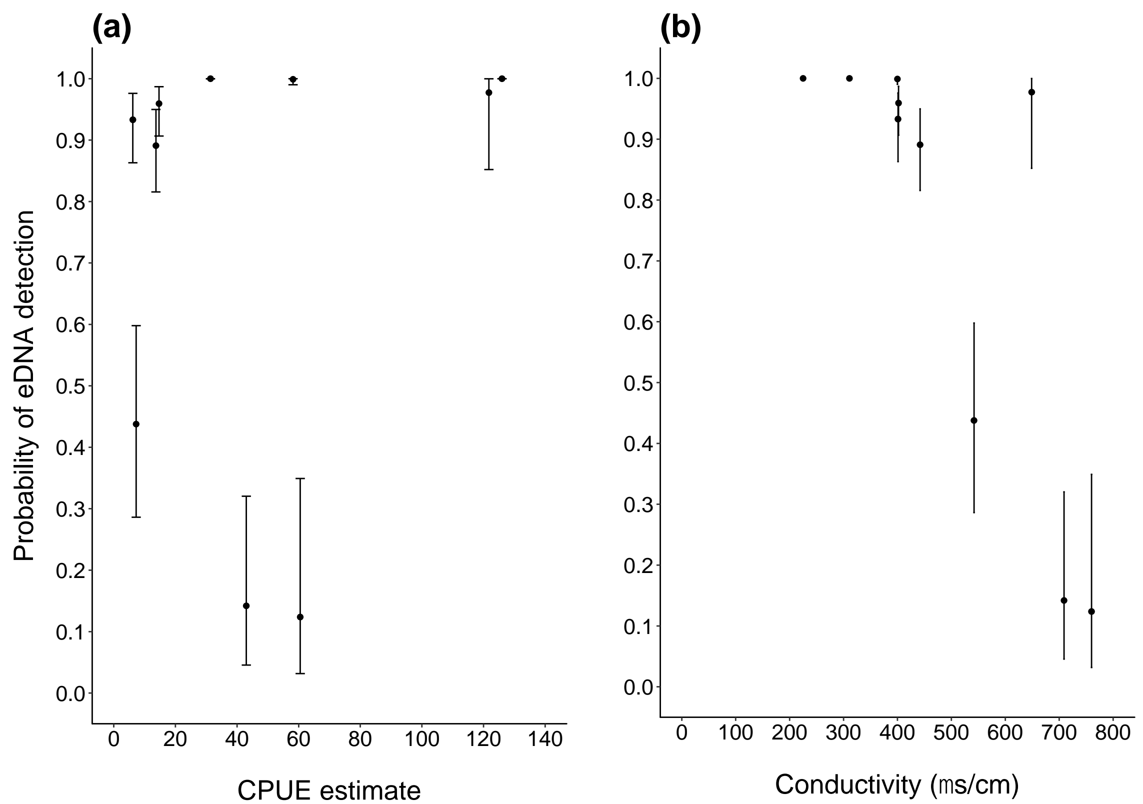


Figure 4.4: Estimated probability of eDNA detection in qPCR replicates. Points are estimates of posterior medians with 95% credible intervals. Probability of eDNA detection in qPCR replicates increased with higher catch-per-unit-effort (CPUE) estimate **(a)** but decreased as conductivity increased **(b)**.

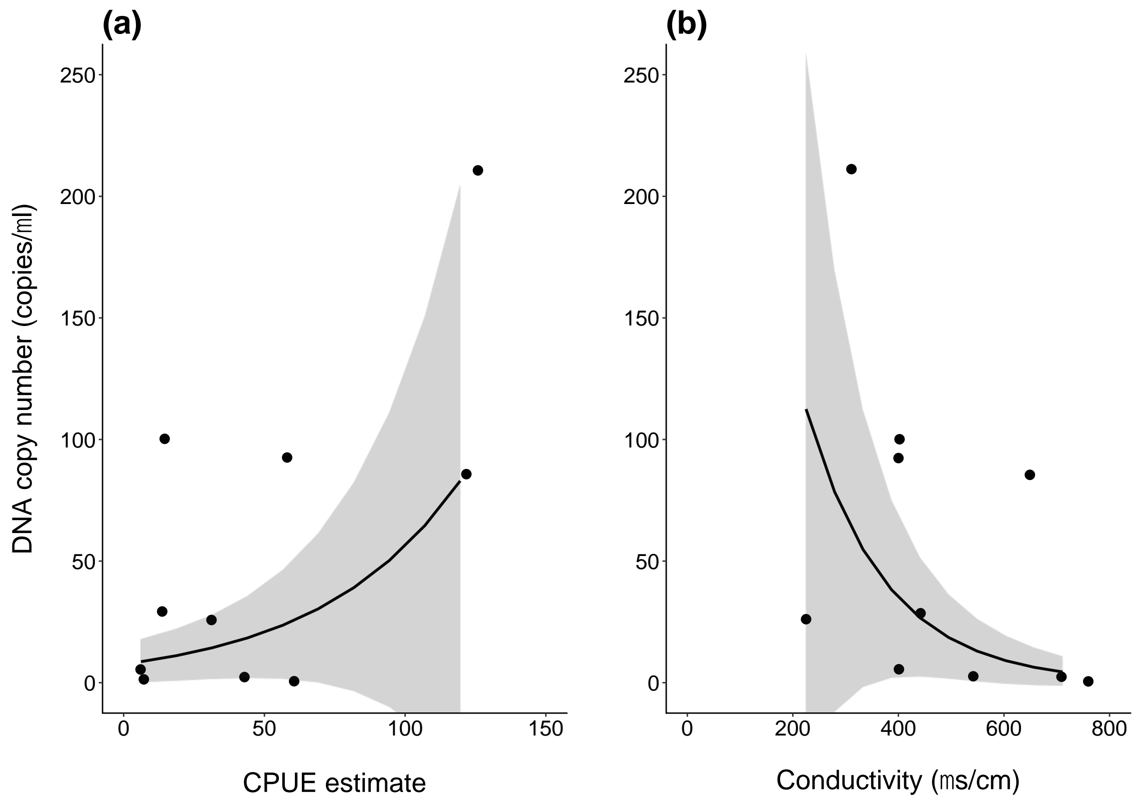


Figure 4.5: Relationship between fixed effects and response variable (DNA copy number) in ponds, as predicted by the Poisson GLMM. The 95% CIs, as calculated using the model predictions and standard error for these predictions, are given for each relationship. The observed data (points) are also displayed against the predicted relationships (line). DNA copy number increased with catch-per-unit-effort (CPUE) estimate **(a)**, but decreased as conductivity **(b)** increased.

4.4 Discussion

We developed a novel species-specific qPCR assay to enable large-scale distribution monitoring of the threatened *C. carassius* using eDNA. *C. carassius* was detected at almost all sites with confirmed presence and no false positives were generated. Furthermore, biotic and abiotic factors that influence eDNA detection and quantification were identified. We discuss areas for improvement in our workflow and provide recommendations for future study.

4.4.1 Using eDNA analysis for *C. carassius* conservation

eDNA analysis is often compared to conventional monitoring tools to assess performance, reliability, reproducibility, and prospective applications in conservation programmes. We found strong agreement between eDNA analysis and fyke netting for *C. carassius* detection, where eDNA analysis detected *C. carassius* in 90% of ponds with presence confirmed by netting. This high detection and low false negative rate supports the applicability of eDNA analysis to *C. carassius* presence-absence monitoring, particularly at large spatial scales where fyke netting can be costly and time-consuming, and where ponds are remote with poor access. Abundance estimation is less straightforward as there was uncertainty around the relationship between DNA copy number and *C. carassius* density. This inconsistency is more likely to result from eDNA analysis than fyke netting due to effects exerted by external factors (section 4.4.2) and sample processing (section 4.4.3) on eDNA quality. However, fyke netting also has detection biases that may influence performance comparisons with eDNA analysis. Fyke net surveys are restricted spatially and temporally to pre- and post-spawning as well as spring and autumn when temperatures are low to reduce fish stress in nets. Furthermore, fyke net surveys may fail to capture species that do not have homogenous distribution in their environment, especially where populations contain few individuals (Turner *et al.*, 2012). Netting is also biased towards particular fish size classes that can enter nets through standard European otter (*Lutra lutra*) guards (75 mm in UK), and catchability is further dependent on time of year (Ruane, Davenport & Igoe, 2012) and even time of day (Hardie, Barmuta & White, 2006). Therefore, effectiveness of standard methods must also be evaluated and eDNA analysis compared to multiple tools before deemed capable or incapable of estimating abundance.

4.4.2 Factors influencing eDNA detection and quantification

Effects of biotic and abiotic factors on eDNA may vary across target species and ecosystems (Barnes *et al.*, 2014). We found *C. carassius* density (CPUE) positively influenced eDNA detection probability and DNA copy number. Density is frequently reported to improve detection probability of aquatic species due to more eDNA deposition in the environment (Schmelzle & Kinziger, 2016; Buxton *et al.*, 2017b; Stoeckle *et al.*, 2017), but this relationship is highly variable across study systems and species due to

influence of external factors (Strickler *et al*., 2015; Buxton *et al*., 2017a; Goldberg *et al*., 2018). For example, increase in water temperature coincided with breeding activity and heightened DNA release in other fish and amphibian species (Buxton *et al*., 2017b; Bylemans *et al*., 2017). In our study, CPUE was collinear with water temperature and thus water temperature was not included in our occupancy model or GLMM. We performed water sample collection in late August, which is outside the reported spawning period for *C. carassius* (Aho & Holopainen, 2000). However, the association between CPUE and DNA copy number may be linked to increased DNA shedding rates caused by higher metabolic activity in response to warm temperature, as reported for other fish species (Takahara *et al*., 2012; Lacoursière-Roussel *et al*., 2016b; Robson *et al*., 2016).

In contrast to CPUE, conductivity had a negative effect on eDNA detection and concentration. Conductivity has been suggested to influence eDNA detection and quantification, but studies that directly measured this variable have found no discernable effect (Takahara *et al*., 2012; Keskin, 2014; Goldberg *et al*., 2018). Conductivity (also measured as TDS) relates to sediment type which influences eDNA detection probability, the rate at which sediment binds eDNA, and release of inhibitory substances (Buxton *et al*., 2017a; Stoeckle *et al*., 2017). Notably, the only false negative pond in our study was also the most conductive (760 µs/cm) and possessed dense beds of rigid hornwort (*Ceratophyllum demersum*) that could restrict water movement. Therefore, conductivity may lead to incorrect inferences about species presence and impact conservation management decisions. Further investigation into the effects of conductivity on eDNA detection and quantification is clearly needed.

Our results indicate that samples may have been affected by inhibitory substances despite tests performed to identify inhibition. We spiked qPCR reactions with a known amount of synthetic target DNA; however, an artificial Internal Positive Control gene assay may identify inhibition more effectively (Goldberg *et al*., 2016). Dilution of eDNA samples (and inhibitory substances present) can release inhibition, but also reduce detection probability (Piggott, 2016) and induce false negatives (Buxton *et al*., 2017a). We used TaqMan® Environmental Master Mix 2.0 (Life Technologies, CA, USA) in qPCR reactions to counter inhibition (Jane *et al*., 2015), but it may be advisable to use DNA extraction kits that perform inhibitor removal (Sellers *et al*., 2018) or include Bovine-serum albumin (BSA) in qPCR reactions (Jane *et al*., 2015). Alternatively, ddPCR may handle inhibitors better than qPCR and provide more accurate abundance or biomass estimates (Nathan *et al*., 2014).

Crucially, environmental data were not collected in 2016 for every pond in our study. Our results indicate direction of effects of biotic and abiotic factors on eDNA detection and quantification, but contemporary data are needed for comprehensive interpretation of these relationships. However, it is clear that eDNA practitioners must account for these effects as well as sample inhibition. The uncertainty around the estimated effects of covariates in our hierarchical occupancy model and GLMM also imply that greater sample volume, sample number, and/or qPCR replication are required to improve the ability and precision of our assay to detect *C. carassius* eDNA and reduce the potential for false negatives (Schultz & Lance, 2015; Goldberg *et al*., 2018).

4.4.3 Optimisation of eDNA workflow

Some non-target DNA extracts used to validate assay specificity were contaminated with *C. carassius* DNA. Field cross-contamination can occur if reference tissue material is collected from multiple species without sterilising equipment, or eDNA is present on the material collected (Rodgers, 2017). Collection and storage of reference tissue material is an important consideration for eDNA practitioners, particularly those using highly sensitive assays (LOD = 1 copy/µL; Wilcox *et al*., 2013, 2016). Dedicated, sterilised equipment should be used when collecting new reference material from different species. From existing reference collections, only non-target samples that were collected on separate and chronologically distinct occasions from target samples should be used (Rodgers, 2017).

Cross-contamination can also arise during water sampling, filtration, DNA extraction and qPCR preparation. Low-level contamination was found in one full process blank but detections from this pond were not omitted as it contained *C. carassius* and contamination was not observed downstream. All equipment in our study was sterilised by immersion in 10% chlorine-based commercial bleach (Elliott Hygiene Ltd, UK) solution for 10 mins, followed by rinsing in 5% MicroSol detergent (Anachem, UK), and then purified water. However, sterilisation with 50% chlorine-based commercial bleach solution (Goldberg *et al*., 2016) or single-use, sterile materials (Wilcox *et al*., 2016) may further minimise contamination risk.

Many of our eDNA samples were low concentration (<100 copies/µL) which can cause inconsistent qPCR amplification (Goldberg *et al*., 2016), thus we discuss approaches to maximise eDNA concentration and improve detection probability. The

probability of eDNA detection depends heavily on the number of samples and volume of water collected, time of sampling, and sample concentration (Schultz & Lance, 2015; Goldberg *et al*., 2018). We sampled 5 x 2 L water samples from each pond in autumn 2016, but timing and/or sampling effort may have been inappropriate. A seasonal effect on *C. carpio* eDNA detection was observed, where spring sampling generated higher eDNA concentration and detection rates due to greater *C. carpio* activity (Turner *et al*., 2014) and density (Hinlo *et al*., 2017a). As water sampling did not coincide with fyke netting (spring 2016) in our study, eDNA concentration may not reflect CPUE estimates. Water samples in spring may contain more *C. carassius* eDNA due to higher activity of individuals, or autumn fyke netting may produce lower CPUE estimates. Parallel seasonal sampling, where water sampling is performed in conjunction with fyke netting at different times of the year, may better align eDNA concentration with CPUE estimates and enable eDNA-based abundance estimates for *C. carassius*. This is certainly a worthwhile area of research.

Representative sampling is crucial in eDNA surveys. Individuals of a species can be unevenly distributed in the environment, which impacts eDNA detection, distribution, and concentration (Takahara *et al*., 2012; Eichmiller, Bajer & Sorensen, 2014; Schmelzle & Kinziger, 2016; Goldberg *et al*., 2018). In lentic ecosystems, eDNA has a patchy horizontal and sometimes vertical distribution, resulting in fine spatial variation (Eichmiller *et al*., 2014). Studies on *C. carpio* revealed eDNA was more concentrated near the shoreline of lentic water bodies (Takahara *et al*., 2012; Eichmiller *et al*., 2014), due to aggregations of individuals (Eichmiller *et al*., 2014). We collected surface water (5 x 2 L) from the shoreline and sampled at equidistant points around the pond perimeter where possible; however, more samples and greater water volumes may be required to improve detection probability (Schultz & Lance, 2015; Goldberg *et al*., 2018). Fine spatial sampling and occupancy modelling are needed to determine the sample number required to achieve high detection probability and eliminate false negatives (Goldberg *et al*., 2018). However, the number of samples required will inevitably vary across habitats due to inherently variable physical properties (Schmelzle & Kinziger, 2016).

Method of eDNA capture can dictate success of this monitoring tool. Studies of eDNA in ponds (Ficetola *et al*., 2008; Biggs *et al*., 2015) have used an ethanol precipitation approach, but this is restricted to small volumes. Filtration allows more water to be processed and minimises capture of non-target DNA, with macro-organism eDNA effectively captured by pore sizes of 1 - 10 μm (Turner *et al*., 2014). We used a

small pore size of 0.45 μm to capture most eDNA particle sizes, although filter clogging prevented the entire sample being processed and may have affected eDNA concentration downstream. Pre-filtering can reduce clogging, but is labour-intensive and increases cost (Takahara *et al*., 2012). Larger pore sizes have been used in temperate and tropical lentic environments (Takahara *et al*., 2012; Robson *et al*., 2016; Goldberg *et al*., 2018), though independent investigation is needed to determine which pore size maximises target DNA concentration.

Comparisons of eDNA yield across filter types and DNA extraction protocols have shown that cellulose nitrate filters stored at -20 °C (this study) often provide best eDNA yield (Piggott, 2016; Spens *et al*., 2016; Hinlo *et al*., 2017b). However, different filter types may be optimal for different species, which has consequences for detectability (Spens *et al*., 2016) and relationships between eDNA concentration and abundance/biomass (Lacoursière-Roussel *et al*., 2016b). Extraction method used, regardless of filter type, will ultimately influence DNA quality and yield. We used the PowerWater® DNA Isolation Kit (MO BIO Laboratories, CA, USA), but the DNeasy Blood and Tissue kit (Qiagen®, Hilden, Germany) has demonstrated greater yield (Hinlo *et al*., 2017b). We also combined filters from the same sample for DNA extraction at the bead milling stage, but independent lysis may recover more DNA (Hinlo *et al*., 2017b). A comparison of DNA extraction protocols is necessary to assess which approach maximises *C. carassius* eDNA concentration. A new modular extraction method shows promise for eDNA but has yet to be evaluated for targeted qPCR (Sellers *et al*., 2018).

Finally, detection sensitivity can be enhanced by increasing the number of qPCR technical replicates (Schultz & Lance, 2015; Piggott, 2016). We performed five technical replicates for each of our samples, but other studies have used as many as twelve and only one may amplify (Biggs *et al*., 2015). More replication may have enabled amplification from the CHIP pond samples, but qPCR cost would inevitably increase. Further replication may also be unnecessary if steps are taken to improve initial concentration of samples instead (Schultz & Lance, 2015).

4.4.4 Concluding remarks

A primary objective of the Norfolk *C. carassius* BAP was to obtain a basic understanding of species distribution and population status across Norfolk (Copp & Sayer, 2010). eDNA surveillance for *C. carassius* will provide a useful, cost-effective alternative to established

survey methods where the aim is determining presence-absence. Our assay may detect hybrids where *C. carassius* were the maternal parent due to use of a mitochondrial marker; however, these detections are also beneficial to the *C. carassius* conservation effort through the identification of ponds where true *C. carassius* may still exist, and where contamination with *C. auratus*, *C. carpio* and their hybrids has occurred. Alternatively, our assay could be used as an early warning tool in countries where *C. carassius* is considered invasive. The areas we have highlighted require further investigation before eDNA analysis can be used routinely. Nevertheless, eDNA survey could enable large-scale distribution monitoring for *C. carassius* through rapid screening of existing and new ponds. Fyke netting could then be used to investigate age, sex and size structure of populations, and remove hybrids.

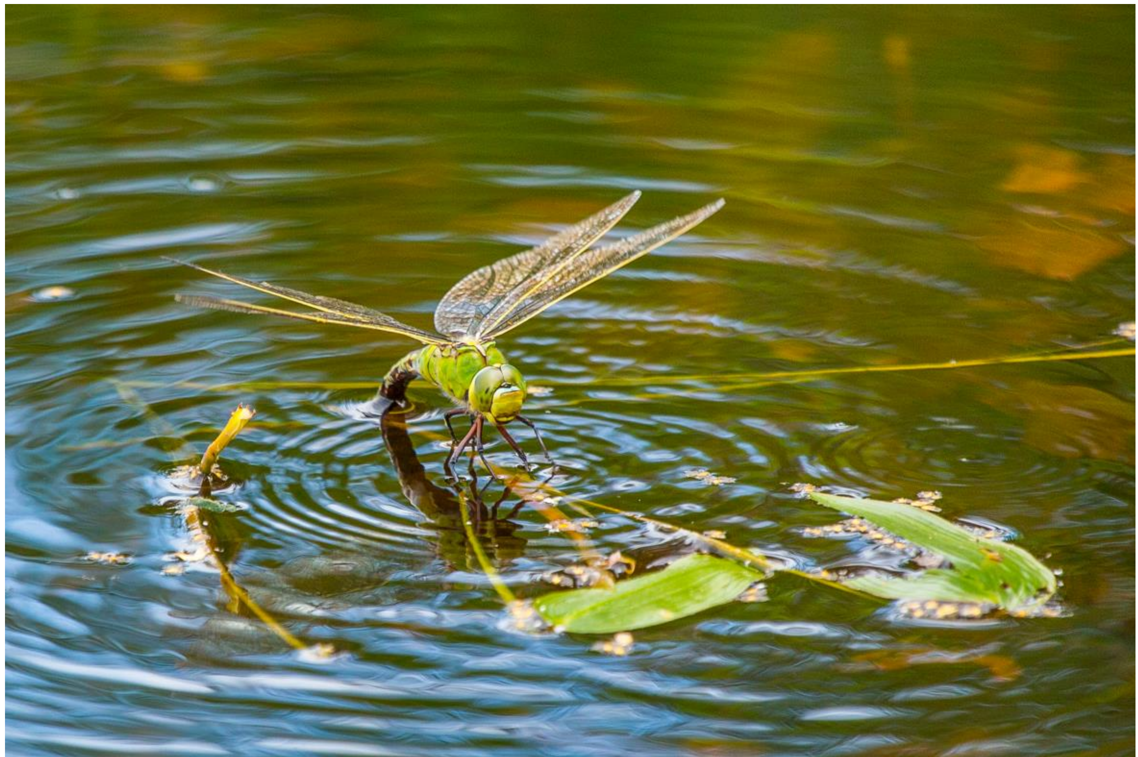
4.5 Acknowledgements

This research was funded by the University of Hull. We would like to thank landowners and land managers for granting permission to sample the ponds included in this study, and Ian Patmore, Dave Emson, Helen Greaves and Glenn Wiseman for assistance with fyke netting. We are grateful to Marco Benucci for assistance with water sampling and filtration, Graham Sellers for advice on primer design and validation, and Peter Shum for support with qPCR troubleshooting and feedback on the manuscript.

4.6 Data accessibility

All R scripts and corresponding data have been deposited in a dedicated GitHub repository (https://github.com/lrharper1/crucian_carp_eDNA_surveillance) which has been permanently archived (https://doi.org/10.5281/zenodo.1421602).

Chapter 5: Assessing the impact of the threatened crucian carp (*Carassius carassius*) on pond invertebrate diversity – a comparison of conventional and molecular tools



Emperor dragonfly (*Anax imperator*) (Leach, 1815)

© user: Derek Parker | Flickr | CC BY-NC-ND 2.0

Abstract

Fish species stocked for recreation and angling can damage freshwater habitats and negatively impact biodiversity, but this is not always the case. The crucian carp (*Carassius carassius*) is one of few fishes naturally associated with ponds and stocked for conservation management. This species may augment landscape-scale diversity; however, its impact on other pond biota has not been broadly assessed. Freshwater invertebrates comprise a large proportion of aquatic diversity, encompassing many rare and endemic species, but are difficult and time-consuming to assess due to small size and high abundance. Practitioners have typically employed sweep-netting and kick-sampling in conjunction with expert morphotaxonomic identification, but DNA and eDNA metabarcoding now provide alternate means to assess invertebrate diversity. These DNA-based approaches can be highly cost-effective and resolve problematic taxa for morphotaxonomic identification. We compared invertebrate diversity in ponds (N = 18) with and without *C. carassius* using sweep-netting and microscopy, DNA metabarcoding, and eDNA metabarcoding. Five 2 L water samples and 4 min sweep-net samples were collected at each pond. Netted samples were identified to lowest taxonomic level possible by generalist microscopy, and these inventories compared to DNA metabarcoding of bulk tissue samples and eDNA metabarcoding of water samples. *C. carassius* presence minimally reduced alpha diversity in ponds, but positively influenced overall beta diversity across ponds through species and family turnover. Ponds with *C. carassius* contained different invertebrate species and families to ponds without fish, resulting in statistically different community composition. eDNA metabarcoding generated the highest alpha diversity, followed by DNA metabarcoding then sweep-netting and microscopy. DNA metabarcoding reflected sweep-netting and microscopy as opposed to eDNA metabarcoding, which produced markedly different communities. Importantly, very few species and families were shared by all three methods, emphasising their complementarity. Therefore, these tools must be used in combination for comprehensive assessment of invertebrate diversity in freshwater ecosystems. Our results will guide pond management in relation to conserving *C. carassius* alongside other biodiversity, and freshwater invertebrate assessment using molecular tools.

5.1 Introduction

Freshwater ecosystems comprise <1% of the Earth's surface, but represent major biodiversity hotspots and provide vital ecosystem services (Dudgeon *et al.*, 2006). Ponds especially provide critical habitat for biodiversity in a fragmented landscape (Céréghino *et al.*, 2008), supporting many rare, unique or specialist species not found in other water bodies (Wood *et al.*, 2003; Biggs *et al.*, 2016). These highly diverse and species-rich ecosystems contribute more to regional-scale diversity than other freshwater habitats, due to their broad-ranging physicochemical properties and greater degree of isolation (Williams *et al.*, 2003; Davies *et al.*, 2008). Aquatic invertebrates are a crucial and abundant component of this diversity, and occupy the vast range of ecological niches made available in ponds by their physicochemical heterogeneity (Williams *et al.*, 2003; Wood *et al.*, 2003; Davies *et al.*, 2008).

In the UK, ponds are commonly stocked with fish for angling and recreation, despite the potentially negative effects on other species, e.g. invertebrates, amphibians (Wood *et al.*, 2001; Gledhill *et al.*, 2008). Fish can alter community structure (Wood *et al.*, 2001; Schilling, Loftin & Huryn, 2009a, b), reduce diversity (Wood *et al.*, 2001; Lemmens *et al.*, 2013), and reduce abundance and biomass (Marklund *et al.*, 2002; Schilling *et al.*, 2009a) of invertebrates. These effects may manifest through direct predation by fish, altered water quality and loss of macrophyte diversity via foraging activity of fish, or management practices associated with angling activity (Wood *et al.*, 2001; Schilling *et al.*, 2009a; Lemmens *et al.*, 2013; Maceda-Veiga *et al.*, 2017). However, the impact of fish stocking can be negligible or even beneficial to invertebrate diversity, particularly at regional-scale, provided that fish species are carefully selected and managed (Gee *et al.*, 1997; Hassall, Hollinshead & Hull, 2011; Lemmens *et al.*, 2013; Stefanoudis *et al.*, 2017).

The crucian carp (*Carassius carassius*) is one of few fish species associated with small ponds in the UK, but has suffered heavy declines and local extinctions in the last century (Copp & Sayer, 2010; Sayer *et al.*, 2011) due to habitat loss, species displacement by the invasive gibel carp (*Carassius gibelio*) (Copp *et al.*, 2008b; Sayer *et al.*, 2011), and genetic introgression through hybridisation (Hänfling *et al.*, 2005). In 2010, *C. carassius* was designated as a Biodiversity Action Plan (BAP) species in the county of Norfolk, England (Copp & Sayer, 2010). A key objective of this BAP is to increase the number of viable *C. carassius* populations across Norfolk through pond restoration and

species reintroduction. Many Norfolk ponds have since been stocked with *C. carassius* to realise this objective (Environment Agency, 2003), but continued stocking is controversial in light of genetic evidence that indicates *C. carassius* is not native to the UK (Jeffries *et al.*, 2017). Nonetheless, there is support for UK conservation efforts to continue to protect the genetic integrity of *C. carassius* at the European level and provide a natural stronghold for the species (Jeffries *et al.*, 2017; Stefanoudis *et al.*, 2017) in the face of persistent declines throughout its native range of Northwest and Central Europe (Copp *et al.*, 2008b; Sayer *et al.*, 2011).

The impact of stocking *C. carassius* on lentic biodiversity has not been thoroughly assessed, and little is known about interactions between this benthic fish and other pond species. Existing research suggests *C. carassius* is characteristic of ponds rich in invertebrates with extensive macrophyte cover (Copp *et al.*, 2008b; Sayer *et al.*, 2011), and plays an important ecological role by increasing landscape-scale diversity across pond networks (Stefanoudis *et al.*, 2017). Yet to our knowledge, only one study has assessed biodiversity (specifically macrophytes, zooplankton, and water beetles) in ponds with *C. carassius*, among other fishes, and without fish (Stefanoudis *et al.*, 2017). Consequently, there is a need to survey and compare fishless ponds to ponds stocked with *C. carassius* to assess the impact of this species on invertebrate diversity more broadly. However, obtaining species resolution data for invertebrates is complicated by the level of taxonomic expertise needed for accurate morphotaxonomic identification as well as cost and time required to survey single sites (Briers & Biggs, 2003; Haase *et al.*, 2010; Hill *et al.*, 2018).

Metabarcoding potentially offers a rapid, high-resolution, cost-effective approach to biodiversity assessment, where multiple species can be identified using High-Throughput Sequencing (HTS) in conjunction with community DNA from bulk tissue samples (DNA metabarcoding), or environmental DNA (eDNA) from environmental samples (eDNA metabarcoding), such as soil or water (Taberlet *et al.*, 2012; Deiner *et al.*, 2017). DNA metabarcoding of aquatic invertebrate samples has proven relatively successful, with applications in biomonitoring and trophic ecology (Andújar *et al.*, 2017; Elbrecht *et al.*, 2017b; Emilson *et al.*, 2017; Lobo *et al.*, 2017; Trevelline *et al.*, 2018). However, only a handful of studies have employed eDNA metabarcoding for invertebrate assessment in freshwater rivers (Deiner *et al.*, 2016; Blackman *et al.*, 2017; Klymus *et al.*, 2017a; Carew *et al.*, 2018b), streams (Macher *et al.*, 2018), and lakes (Klymus *et al.*, 2017a). To date, there are no published studies that have used metabarcoding for pond

invertebrates.

We assessed invertebrate diversity in ponds with and without *C. carassius* using metabarcoding in conjunction with standard sweep-net surveys and morphotaxonomic identification. The effect of *C. carassius* stocking on invertebrate diversity was determined using the species inventories generated by each tool, both individually and combined. These inventories were then compared to evaluate which monitoring tool provides the most holistic assessment of invertebrate diversity. We hypothesised that alpha diversity would be lower in ponds with *C. carassius*, but beta diversity would be enhanced due to heterogeneity induced by *C. carassius* across the pond network. This pattern was expected regardless of monitoring tool used. We anticipated that DNA metabarcoding and morphotaxonomic identification would produce complementary views of pond invertebrate communities, whereas eDNA metabarcoding would reveal species not identified by DNA metabarcoding or microscopy. DNA and eDNA metabarcoding were expected to enable species resolution for some problematic taxa that cannot be morphologically identified to species-level using standard keys. We provide recommendations for pond management, specifically conservation of *C. carassius* alongside pond biodiversity, and the application of molecular tools to freshwater invertebrate assessment.

5.2 Materials and methods

5.2.1 Study sites

We surveyed nine ponds with confirmed *C. carassius* presence at different densities across Norfolk and East Yorkshire, and nine fishless ponds in Norfolk. All study ponds were selected to be <1 ha in area, <5 m in depth, macrophyte-dominated, with a largely open canopy and thus minimal shading of the water surface. Ponds were mainly surrounded by arable fields, excluding one located in woodland. No specific permits were required for sampling, but relevant landowner permissions were obtained. Samples for morphotaxonomic identification and metabarcoding were all collected in autumn 2016 at peak invertebrate diversity (Hill, Sayer & Wood, 2016). Data on physical (area, depth, percentages of perimeter with emergent vegetation, emergent macrophyte cover, submerged macrophyte cover, and shading) and chemical (conductivity) properties of

ponds were collected between May and August from 2010 to 2017. Conductivity was measured with a HACH HQ30d meter (Hach Company, CO, USA). Percentages of perimeter with emergent vegetation, emergent macrophyte cover, submerged macrophyte cover, and shading of ponds by trees and scrub were estimated visually.

5.2.2 Sweep-netting and morphotaxonomic identification

Sweep-netting was performed in accordance with the UK National Pond Survey methodology (Biggs, Fox & Nicolet, 1998), using a standard 1 mm mesh long-handled net (0.3 m square bag), to generate a conventional taxonomic inventory of lentic invertebrates. Sampling time at each pond totalled 4 min, with 3 min of sweep-netting and a 1 min hand search. The time allotted to sweep-netting was divided equally across identified mesohabitats, e.g. emergent macrophytes, submerged macrophytes, shaded water, marginal grasses, open water. In ponds with dominant mesohabitat, sampling time was divided to take additional samples from the dominant mesohabitat. For example, in a pond with 3 mesohabitats (one dominant), sampling time was divided by 4 (Biggs *et al.*, 1998). During the 1 min search, the water surface and hard substrate (e.g. rocks, logs) were inspected for aquatic invertebrates additional to those collected in the net. Collected material from sweeps and searches were pooled to create one sample for each pond, and deposited in a 1.2 L sterile Whirl-Pak® stand-up bag (Cole-Palmer, Hanwell, London). Samples were transported in a sterile coolbox with ice to the University of Hull, and stored at -20 ^{o}C until samples could be processed and sorted in the laboratory. Each sample was thawed and passed through sieves of 8 mm, 2 mm, and 250 mm to remove large items of vegetation and detrital matter. Specimens were identified under a light microscope to family-level (Dobson *et al.* 2012). Terrestrial taxa, empty Trichoptera (caddisfly) cases, and empty shells were discarded. All specimens were preserved in sterile 15 mL falcon tubes (SARSTED, Germany) containing 100% ethanol according to family and pond sampled (N = 18), and stored at -20 ^{o}C until further processing.

An additional five ponds (four with *C. carassius* and one fishless) in Norfolk were sampled for invertebrates as outlined above to obtain specimens representative of different species, families, and major groups for metabarcoding primer validation. Specimens of newly inventoried species were removed for individual preservation in sterile 2 mL microtubes (Fisher Scientific UK Ltd, UK) with 100% ethanol, and stored at -20 ^{o}C until DNA extraction. Each species was extracted individually, using a leg as

starting tissue material, with the DNeasy Blood & Tissue Kit® (Qiagen®, Hilden, Germany) following the manufacturer's protocol. DNA extracts were then stored at at -20 °C until PCR. These five ponds were not included for bulk tissue DNA metabarcoding.

5.2.3 DNA metabarcoding samples

We largely followed the workflows established by Elbrecht, Peinert & Leese (2017a) and Blackman *et al.* (2017) for DNA metabarcoding of invertebrate bulk tissue samples. Specimens were sorted into three body size categories on laminated millimetre graph paper: small (S, below 2.5 x 5 mm), medium (M, 2.5 x 5 mm up to 5 x 10 mm), and large (L, greater than 10 mm and up to 10 x 20 mm). During size-sorting, specimens were identified under a light microscope to species-level where possible, using Freshwater Biological Association publications (Macan, 1960; Friday, 1988; Savage, 1989; Wallace *et al.*, 1990; Gledhill, Sutcliffe & Williams, 1993; Edington & Hildrew, 1995; Bass, 1998; Elliott, 2009; Elliott & Humpesch, 2010; Brooks & Cham, 2014; Elliott & Dobson, 2015). The laminated paper was sterilised with 50% v/v chlorine-based commercial bleach solution (Elliot Hygiene Ltd, UK) and 80% v/v ethanol solution between measuring specimens from different ponds to minimise cross-contamination risk. Specimens were preserved in sterile 15 mL falcon tubes (SARSTED, Germany) containing 100% ethanol according to size category and pond sampled, and stored at -20 °C until DNA extraction.

Size categories from each pond were dried overnight on FisherBrand cellulose filter paper (Fisher Scientific UK Ltd, UK) in sterile glass funnels and conical flasks to remove excess ethanol. Size categories were then lysed (3 × 30 sec) using a Qiagen Tissue Lyser® (Qiagen®, Hilden, Germany) with DigiSol (50mM Tris, 20M EDTA, 120 mM NaCl and 1% SDS). The TissueLyser adapter sets could only hold 1.5 g of dried tissue and corresponding volume of DigiSol. Therefore, if the dry tissue weight of any size category exceeded 1.5 g, we processed the size category in batches until all tissue had been lysed. The lysates from all batches were then pooled to recreate size categories. The size categories were incubated overnight at 55 °C with SDS and Proteinase K (Bioline®, London, UK), following which 200 μL of lysate from each size category was used for extraction with the DNeasy Blood & Tissue Kit® (Qiagen®, Hilden, Germany) according to the manufacturer's protocol. Consequently, 16 bulk tissue samples were represented by three DNA extracts and two bulk tissue samples represented by two DNA extracts (*n*

= 52) that were sequenced individually (see Appendix 5). An extraction blank, consisting only of extraction buffers, was included for each round of DNA extraction.

5.2.4 eDNA metabarcoding samples

eDNA samples used in Chapter 4 for validation of a quantitative PCR assay for *C. carassius* (see also Harper *et al*., 2019b) were repurposed here for eDNA metabarcoding of invertebrate communities. Briefly, five 2 L surface water samples were collected from the shoreline of each pond at equidistant intervals where access permitted. Water samples were transported on ice in sterile coolboxes to the Centre for Ecology & Hydrology, Wallingford, stored at 4 °C, and vacuum-filtered within 24 hours of collection. For each pond, a full process blank (1 L molecular grade water) was taken into the field and stored in coolboxes with samples. Blanks were filtered and extracted alongside samples to identify contamination. DNA was isolated from filters using the PowerWater® DNA Isolation Kit (MO BIO Laboratories, CA, USA) and following the manufacturer's protocol in a dedicated eDNA facility at the University of Hull. This facility is devoted to pre-PCR processes with separate rooms for filtration, DNA extraction, and PCR preparation of environmental samples. DNA extracts were stored at -20 °C until further analysis.

5.2.5 Metabarcoding workflow

A comprehensive list of UK invertebrate species living in or associated with freshwater habitats was established by the Centre for Ecology & Hydrology (see https://www.ceh.ac.uk/services/coded-macroinvertebrates-list). This list was used to create custom curated reference databases for UK aquatic invertebrates, excluding Diptera. Reference databases were constructed from sequences deposited in the public NCBI database GenBank in August 2017 (see Appendix 5). Public records for Diptera were missing record features (e.g. 'gene' or 'CDS') and/or names were not in the format required for custom reference database construction using the selected bioinformatic tools. The extent of reference sequence representation for species varied across the invertebrate databases at time of curation (Fig. S5.1): Coleoptera 97.40% (*N* = 423 species), Odonata 91.53% (*N* = 59), Hemiptera and Hymenoptera 46.49% (*N* = 114), Trichoptera and Lepidoptera 90.78% (*N* = 206), Ephemeroptera, Plecoptera, Neuroptera,

and Megaloptera 90.22% (*N* = 92), Crustacea 39.69% (*N* = 388), Mollusca 70.27% (*N* = 111), Arachnida 100% (*N* = 333), and Annelida 84.87% (*N* = 152). The complete reference databases compiled in GenBank format have been deposited in a dedicated GitHub repository for this chapter, permanently archived at: https://doi.org/10.5281/zenodo.2634240. Species without database representation (Fig. S5.2) are listed in an excel file provided in the archived GitHub repository for this chapter (https://doi.org/10.5281/zenodo.2634240).

Published primers mICOIintF (Leray *et al.*, 2013) and jgHCO2198 (Geller *et al.*, 2013), which amplify a 313 bp fragment of the mitochondrial cytochrome c oxidase subunit I (*COI*) gene, were selected for metabarcoding. The primers were validated *in silico* using ecoPCR software (Ficetola *et al.*, 2010) against the custom invertebrate reference databases. Parameters set allowed a 250-350 bp fragment and maximum of three mismatches between the primer pair and each sequence in the reference database. Primers were then validated *in vitro* for 38 invertebrate species, representing 38 families and 10 major groups (Fig. S5.3). Primer performance was also evaluated *in vitro* against other published metabarcoding primers (Figs. S5.4, S5.5) for macroinvertebrates (Elbrecht & Leese, 2017; Vamos, Elbrecht & Leese, 2017).

After primer validation, PCR conditions were optimised (Figs. S5.6, S5.7) and two independent libraries were constructed for DNA metabarcoding and eDNA metabarcoding using a two-step PCR protocol. During the first PCR, the target region was amplified using metabarcoding primers, comprised of the aforementioned specific locus primer, sequencing primer, and pre-adapter (Illumina, 2011). DNA from the exotic, terrestrial two-spotted assassin bug (*Platymeris biguttatus*) was used for PCR positive controls (tissue DNA *N* = 9, eDNA *N* = 11) as this species is not found in the UK, whilst sterile molecular grade water (Fisher Scientific UK Ltd, UK) substituted template DNA for PCR negative controls (tissue DNA *N* = 9, eDNA *N* = 11). PCR products were individually purified using a magnetic bead clean-up (VWR International Ltd, UK), following a double size selection protocol from Bronner *et al.* (2009). The second PCR bound Multiplex Identification (MID) tags to the purified products. PCR products were pooled according to PCR run and the pooled PCR product purified using a magnetic bead clean-up, following a double size selection protocol from Bronner *et al.* (2009). Each purified PCR product was quantified on a Qubit™ 3.0 fluorometer using a Qubit™ dsDNA HS Assay Kit (Invitrogen, UK) and normalised according to concentration and sample number to produce a pooled volume of 20 μL.

The pooled libraries were quantified on a Qubit™ 3.0 fluorometer and diluted to 6 nM for quantification by real-time quantitative PCR using the NEBNext® Library Quant Kit for Illumina® (New England Biolabs® Inc., MA, USA). The libraries were also checked using an Agilent 2200 TapeStation and High Sensitivity D1000 ScreenTape (Agilent Technologies, CA, USA) to verify all secondary product was removed by bead purification and only a fragment of the expected size (531 bp) remained. A total of 52 bulk tissue subsamples sequenced in triplicate (n = 156), 12 extraction blanks, and 18 PCR controls alongside samples from other projects (N = 188) were included in the bulk tissue library. A total of 90 eDNA samples, 18 full process blanks, and 22 PCR controls alongside samples from other projects (N = 140) were included in the eDNA library. The bulk tissue library with 10% PhiX Sequencing Control and eDNA library with 20% PhiX Sequencing Control were sequenced on an Illumina® MiSeq using 2 x 300 bp V3 chemistry (Illumina, Inc, CA, USA).

Illumina® data was converted from raw sequences to taxonomic assignment using a custom pipeline for reproducible analysis of metabarcoding data: metaBEAT (metaBarcoding and eDNA Analysis Tool) v0.97.11 (https://github.com/HullUni-bioinformatics/metaBEAT). After quality trimming, merging, chimera detection, and clustering, non-redundant query sequences were compared against our reference database using BLAST (Zhang *et al.*, 2000). Putative taxonomic identity was assigned using a lowest common ancestor (LCA) approach based on the top 10% BLAST matches for any query matching with at least 90% identity to a reference sequence across more than 80% of its length. Unassigned sequences were subjected to a separate BLAST against the complete NCBI nucleotide (nt) database at 90% identity to determine the source via LCA as described above. Bioinformatic settings were chosen based on comprehensive exploration of the parameter space and comparison of metaBEAT taxonomic assignments to morphotaxonomic inventories. The bioinformatic analysis has been deposited in the GitHub repository for reproducibility. Full details of the metabarcoding workflow are provided in Appendix 5.

5.2.6 Data analysis

Analyses were performed in the statistical programming environment R v.3.4.3 (R Core Team, 2017). Data and R scripts have been deposited in the GitHub repository. Assignments from the custom and public databases were merged, and spurious

assignments (i.e. non-metazoans) removed from the datasets. Assignments corresponding to ambiguous BOLD records were renamed as the genus or family stated in the record name. Reads from the same assignments were then merged. To minimise risk of false positives, taxa were only classed as present at sites if their sequence frequency exceeded set thresholds. For the DNA metabarcoding dataset, this threshold was defined using the maximum sequence frequency of the PCR positive control (*P. biguttatus* DNA) in the bulk tissue samples (0.00016%). However, there was no *P. biguttatus* contamination of eDNA samples, thus taxon-specific thresholds were applied to the eDNA metabarcoding dataset instead (Harper *et al.*, 2018a). The thresholds were defined using the maximum sequence frequency of each taxa in the PCR positive controls (N = 11). Only *Homo sapiens* (0.00075%) and the unassigned reads (0.0171%) required thresholds.

After applying the false positive thresholds, non-invertebrate assignments and coarse invertebrate assignments (above family-level) were removed. For the DNA metabarcoding dataset, we then pooled the sequence data for PCR/sequencing replicates belonging to the same size category, and pooled size categories according to pond sampled. For the eDNA metabarcoding dataset, we pooled the sequence data for biological replicates belonging to the same pond. Subsets of the DNA and eDNA metabarcoding datasets were created that contained the species-level and family-level assignments respectively for each sample. The abundance (morphotaxonomic identification) and read count data (metabarcoding) were then converted to site x taxonomy presence-absence matrices for downstream analysis using the *decostand* function in the R package vegan v2.4-6 (Oksanen *et al.*, 2018). Presence-absence matrices were used as potential amplification bias during PCR can prevent reliable abundance or biomass estimation from sequence reads produced by DNA or eDNA metabarcoding (Elbrecht *et al.*, 2017a). First, the effect of *C. carassius* on invertebrate diversity was assessed at species-level and family-level according to method used (sweep-netting and microscopy, DNA metabarcoding, eDNA metabarcoding). The data produced by each method were then combined at species-level and family-level respectively, and the impact of *C. carassius* alongside environmental variables on combined invertebrate diversity investigated. Finally, the species and family inventories produced by each method of freshwater invertebrate assessment were compared.

For the purposes of this chapter, we define alpha diversity as the raw taxon richness of ponds, and beta diversity as the difference between communities present at each pond whilst accounting for taxon identity (Baselga & Orme, 2012). For each data

set, the following analyses were performed. Alpha diversity was obtained using the *specnumber* function in vegan v2.4-6 (Oksanen *et al.*, 2018). Total alpha diversity of ponds (response variable) was modelled against *C. carassius* presence-absence (explanatory variable) using a Generalized Linear Model (GLM). Using the combined method data set only, alpha diversity of the major invertebrate groups (response variable), i.e. orders, classes, or phyla (Dobson *et al.*, 2012), was modelled against *C. carassius* presence-absence (explanatory variable) using a GLM. Total alpha diversity of ponds (response variable) was then modelled against sampling method (explanatory variable) using a GLM. A negative binomial distribution was specified for all GLMs. Pairwise Tukey's HSD tests were used to determine whether the differences in alpha diversity were significant. The R package betapart v1.5.0 (Baselga & Orme, 2012) was used to estimate total beta diversity, partitioned by nestedness and turnover, across all ponds and sampling methods with the *beta.multi* function. These three components of beta diversity were then estimated for ponds with or without *C. carassius*, and inventories produced by each sampling method, using the *beta.pair* function. For each component of beta diversity, we compared community heterogeneity in each group of ponds and sampling method by calculating homogeneity of multivariate dispersions (MVDISP) using the *betadisper* function from vegan v2.4-6 (Oksanen *et al.*, 2018). Variation in MVDISP was then statistically tested using an ANOVA, and pairwise Tukey's HSD tests used to determine if there were significant differences between the groups (*C. carassius* presence-absence or sampling method). The effect of *C. carassius* and sampling method on each component of beta diversity was visualised using Non-metric Multidimensional Scaling (NMDS) with the *metaMDS* function, and tested statistically using permutational multivariate analysis of variance (PERMANOVA) with the function *adonis* in vegan v2.4-6 (Oksanen *et al.*, 2018). Pre-defined cut-off values were used for effect size, where PERMANOVA results were interpreted as moderate and strong effects if $R^2 > 0.09$ and $R^2 > 0.25$ respectively. These values are broadly equivalent to correlation coefficients of $r = 0.3$ and 0.5 which represent moderate and strong effects accordingly (Nakagawa & Cuthill, 2007; Macher *et al.*, 2018). Jaccard dissimilarity was used as a measure of beta diversity for all analyses.

We tested whether the invertebrate communities produced by all three sampling methods combined were influenced by the physical and chemical properties of ponds in conjunction with *C. carassius* presence-absence at species-level and family-level. Redundancy Analysis (RDA) was selected for constrained ordination as it analyses

variation in biological communities in response to explanatory variables (Legendre & Legendre, 2012). Principal Coordinate Analysis (PCoA) was performed using the *pcoa* function in the R package ape v5.0 (Paradis & Schliep, 2018) on the turnover, nestedness, and total beta diversity matrices generated for the combined data using *beta.pair* function in vegan v2.4-6 (Oksanen *et al.*, 2018). The Lingoes correction was employed to account for negative eigenvalues (Legendre, 2014). The resultant PCoA eigenvectors (principle coordinates) for each distance matrix were used as the response variable in variance partitioning analysis. Our variables were grouped as biotic (*C. carassius* presence-absence) or abiotic (pond conductivity, area, depth, percentages of perimeter with emergent vegetation, emergent macrophyte cover, submerged macrophyte cover, and shading) for the purposes of RDA and variance partitioning. Abiotic variables were log10 transformed to eliminate their physical units (Legendre & Birks, 2012). Significant abiotic variables influencing each component of beta diversity were identified using the *ordiR2step* function in vegan v2.4-6 to perform separate RDA analyses under a forward selection procedure (Oksanen *et al.*, 2018). Where applicable, the relative contributions of the biotic and abiotic variables on turnover, nestedness, and total beta diversity for the species-level and family-level invertebrate communities were then assessed by variance partitioning (Borcard, Legendre & Drapeau, 1992) using the *varpart* function from vegan v2.4-6 (Oksanen *et al.*, 2018). For each beta diversity component, RDA was performed using our biotic and identified significant abiotic variables, and variance partitioning used to divide the total percentage of variation explained into unique and shared contributions for biotic and abiotic predictor groups. The *anova* function in vegan v2.4-6 (Oksanen *et al.*, 2018) was used to examine the statistical significance of the full model and the unique contributions of each predictor group. We report the adjusted R^2-fractions in this study as they are widely recommended and unbiased (Peres-Neto *et al.*, 2006).

5.3 Results

5.3.1 Taxonomic composition by method

5.3.1.1 Sweep-netting and morphotaxonomic identification

Across samples from 18 ponds, we identified 2,281 specimens belonging to 38 families, and from this total, 1,404 specimens were identified as belonging to 91 species (see lists provided in archived GitHub repository: https://doi.org/10.5281/zenodo.2634240). Overall, the most abundant taxa were *Asellus aquaticus* (11.68%), *Pisidium casertanum* (7.69%), *Erpobdella octoculata* (5.91%), *Coenagrion puella* (5.41%), and *Radix peregra* (5.27%) at species-level, and Chironomidae (33.14%), Asellidae (8.24%), and Coenagrionidae (8.20%) at family-level. However, *Notonecta glauca* (*n* = 12 ponds), *C. puella* (*n* = 10), and *Enallagma cyathigerum* (*n* = 10) occurred in the most ponds at species-level, and Dytiscidae (*n* = 17), Chironomidae (*n* = 16), and Coenagrionidae (*n* = 15) occurred in the most ponds at family-level.

5.3.1.2 DNA metabarcoding

The sequencing run generated 34,473,112 raw sequence reads. In total, 12,024,697 sequences remained after trimming, merging, chimera removal, and clustering (average read count of 32,324 per sample). From these sequences, 7,281,801 (60.56%) were assigned to a metazoan or non-metazoan taxonomic rank, but 4,742,896 were not assigned a taxonomic identity (39.44%). Across the study ponds, 2,454,295 and 2,907,165 sequence reads were assigned to 141 species and 57 families respectively (see lists provided in archived GitHub repository: https://doi.org/10.5281/zenodo.2634240) after assignments were corrected (i.e. removed, renamed, or merged), application of the false positive sequence threshold, and removal of coarse assignments and samples from other projects. The majority of reads were assigned to *N. glauca* (14.82%), *A. aquaticus* (8.20%), *E. octoculata* (5.88%), and *Chironomus luridus* (5.28%) at species-level, and Notonectidae (14.60%), Chironomidae (13.06%), and Corixidae (9.33%) at family-level. The taxa that inhabited the most ponds were *N. glauca* (*n* = 16 ponds), *C. puella* (*n* = 16), *A. aquaticus* (*n* = 12), and *C. luridus* (*n* = 12) at species-level, and Chironomidae (*n* = 18), Naididae (*n* = 17), Coenagrionidae (*n* = 16), Dytiscidae (*n* = 16), Notonectidae (*n* =

16), and Corixidae (*n* = 15) at family-level.

5.3.1.3 eDNA metabarcoding

The sequencing run generated 11,019,530 raw sequence reads. In total, 4,267,530 sequences remained after trimming, merging, chimera removal, and clustering (average read count of 16,075 per sample). From these sequences, 1,726,801 (40.46%) were assigned to a metazoan or non-metazoan taxonomic rank, but 2,540,729 were not assigned a taxonomic identity (59.54%). Across the study ponds, 389,766 and 831,073 sequence reads were assigned to 160 species and 92 families respectively (see lists provided in archived GitHub repository: https://doi.org/10.5281/zenodo.2634240) after assignments were corrected (i.e. removed, renamed, or merged), application of the taxon-specific false positive sequence thresholds, and removal of coarse assignments and samples from other projects. The majority of reads were assigned to *Cyclops strenuus* (12.21%), *Cloeon dipterum* (11.49%), and *Keratella cochlearis* (10.38%) at species-level, and Cyclopidae (41.52%), Chironomidae (14.94%), Brachionidae (6.67%), Naididae (6.58%), and Baetidae (5.39%) at family-level. However, the most common taxa across the study ponds were *Rotaria rotatoria* (*n* = 16 ponds), *Chaetogaster diastrophus* (*n* = 14), *C. dipterum* (*n* = 14), and *Eucyclops serrulatus* (*n* = 14) at species-level, and Chironomidae (*n* = 18), Cyclopidae (*n* = 18), Macrotrichidae (*n* = 17), and Philodinidae (*n* = 17) at family-level.

5.3.1.4 Combined methods

The three methods of invertebrate assessment combined identified 392 species and 187 families across the study ponds (Tables S5.1, S5.2). The combined data indicated that *C. puella* (*n* = 16 ponds), *N. glauca* (*n* = 16), *R. rotatoria* (*n* = 16), *C. dipterum* (*n* = 15), and *C. diastrophus* (*n* = 15) were the most common species, and Chironomidae (*n* = 18), Dytiscidae (*n* = 18), Cyclopidae (*n* = 18), and Naididae (*n* = 18) were the most common families.

5.3.2 Impact of *C. carassius* stocking on pond invertebrates

Independently and combined, methods revealed overall alpha diversity of invertebrates was marginally reduced in ponds containing *C. carassius* at species-level (Fig. 5.1ai-iv) and family-level (Fig. 5.1bi-iv), but these differences were not significant (Table 5.1). Detailed examination of alpha diversity within the major invertebrate groups (Dobson *et al.*, 2012) identified by all three methods combined revealed that Coleoptera and Mollusca diversity was significantly reduced in ponds with *C. carassius* at species-level (GLM: Coleoptera -0.534 ± -0.255, Z = -2.095, P = 0.036; Mollusca -0.815 ± 0.268, Z = -3.043, P = 0.002), but not family-level (GLM: Coleoptera -0.511 ± -0.298, Z = -1.713, P = 0.087; Mollusca -0.442 ± -0.302, Z = -1.462, P = 0.144). However, differences in alpha diversity between ponds with or without *C. carassius* were not significant for other invertebrate groups at either taxonomic rank (Fig. 5.2).

Table 5.1: Summary of analyses (GLM) statistically comparing alpha diversity (taxon richness) at species-level and family-level between ponds with and without *C. carassius* using independent and combined methods.

	Generalized Linear Model (GLM)							
	Species-level				**Family-level**			
	df	**Estimate (SE)**	***Z***	***P***	**df**	**Estimate (SE)**	***Z***	***P***
Sweep-netting and microscopy	1	-0.139 (0.186)	-0.746	0.456	1	-0.098 (0.158)	-0.620	0.535
DNA metabarcoding	1	0.046 (0.148)	0.312	0.755	1	-0.015 (0.120)	-0.120	0.904
eDNA metabarcoding	1	-0.045 (0.198)	-0.227	0.820	1	0.147 (0.141)	-1.043	0.297
Combined methods	1	-0.049 (0.122)	-0.403	0.687	1	-0.104 (0.097)	-1.067	0.286

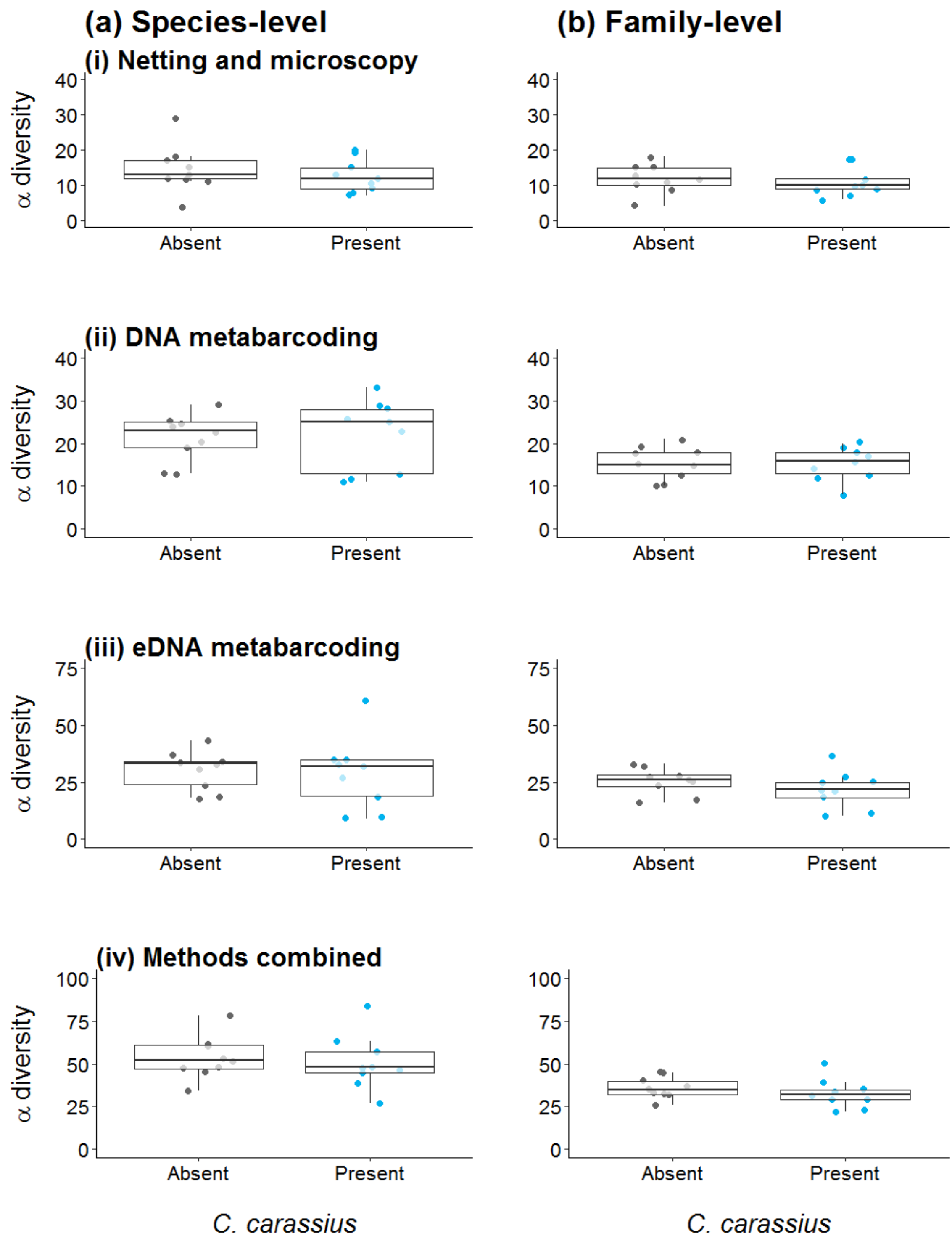


Figure 5.1: Mean alpha diversity (taxon richness) of invertebrates in ponds with *C. carassius* (blue points) and without fish (grey points) across Norfolk and East Yorkshire. Alpha diversity at species-level **(a)** and family-level **(b)** is shown according to method of invertebrate assessment: sweep-netting and microscopy **(i)**, DNA metabarcoding **(ii)**, eDNA metabarcoding **(iii)**, and all methods combined **(iv)**. Boxes show 25th, 50th, and 75th percentiles, and whiskers show 5th and 95th percentiles.

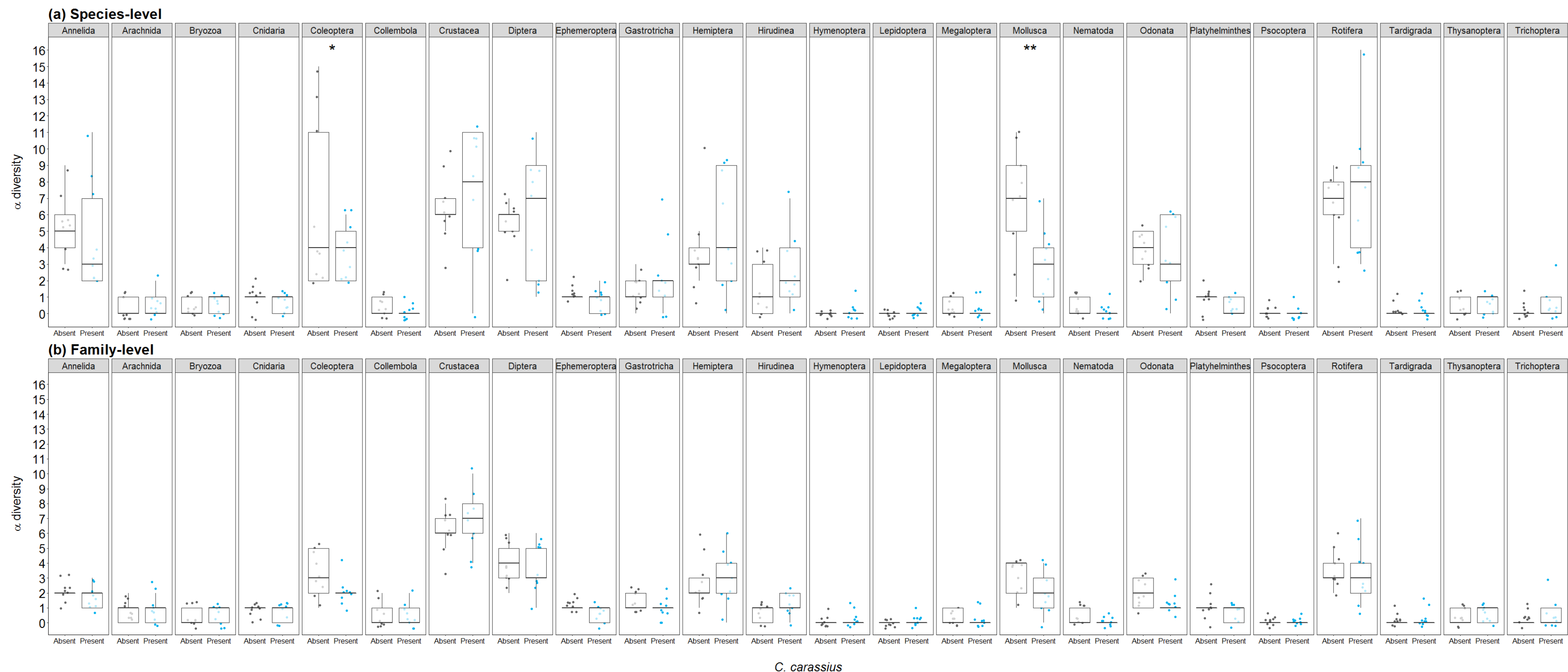


Figure 5.2: Mean alpha diversity (taxon richness) at species-level (a) and family-level (b) of the different invertebrate groups identified by all three survey methods in ponds with *C. carassius* (blue points) and without fish (grey points) across Norfolk and East Yorkshire. Boxes show 25th, 50th, and 75th percentiles, and whiskers show 5th and 95th percentiles. Significant differences are indicated by asterisks (* = $P < 0.05$, ** = $P < 0.01$).

Beta diversity was comparable for independent and combined methods, where total beta diversity of ponds was consistently high at species-level and family-level. Variation in invertebrate community composition was predominantly driven by turnover rather than nestedness (Table 5.2). Using either sweep-netting and microscopy or DNA metabarcoding, homogeneity of multivariate dispersions (MVDISP) was not significantly different between ponds for turnover or total beta diversity; although ponds with *C. carassius* had significantly lower dispersion than ponds without fish for nestedness. Conversely, MVDISP was not significantly different between ponds for turnover or nestedness using eDNA metabarcoding or methods combined. Instead, ponds without fish had significantly lower dispersion than ponds with *C. carassius* for total beta diversity (Table 5.3).

Table 5.2: Relative contribution of species turnover and nestedness to total beta diversity (Jaccard dissimilarity). A value of 1 corresponds to all sites containing different species.

	Species-level			**Family-level**		
	Turnover	**Nestedness**	**Total beta diversity**	**Turnover**	**Nestedness**	**Total beta diversity**
Sweep-netting and microscopy	0.935 (98.01%)	0.019 (1.99%)	0.954 (100%)	0.867 (94.75%)	0.048 (5.25%)	0.915 (100%)
DNA metabarcoding	0.938 (98.53%)	0.014 (1.47%)	0.952 (100%)	0.883 (96.61%)	0.031 (3.39%)	0.914 (100%)
eDNA metabarcoding	0.917 (97.24%)	0.026 (2.76%)	0.943 (100%)	0.870 (95.29%)	0.043 (4.71%)	0.913 (100%)
Combined methods	0.927 (98.41%)	0.015 (1.59%)	0.942 (100%)	0.868 (96.23%)	0.034 (3.77%)	0.902 (100%)

Table 5.3: Summary of analyses (ANOVA) statistically comparing homogeneity of multivariate dispersions (MVDISP) between the communities in ponds with and without *C. carassius* as well as the communities produced by each method of invertebrate assessment at species-level and family-level.

	Homogeneity of multivariate dispersions (ANOVA)							
	Species-level				**Family-level**			
	Mean distance to centroid ± SE	**df**	***F***	***P***	**Mean distance to centroid ± SE**	**df**	***F***	***P***
Netting and microscopy								
Turnover		1	3.706	0.072		1	1.522	0.235
C. carassius	0.562 ± 0.002				0.393 ± 0.009			
No *C. carassius*	0.502 ± 0.007				0.331 ± 0.014			
Nestedness		1	5.090	**0.038**		1	2.908	0.108
C. carassius	0.035 ± 0.001				0.087 ± 0.002			
No *C. carassius*	0.069 ± 0.002				0.141 ± 0.007			
Total beta diversity		1	1.576	0.227		1	0.744	0.401
C. carassius	0.588 ± 0.001				0.468 ± 0.007			
No *C. carassius*	0.554 ± 0.006				0.428 ± 0.013			
DNA metabarcoding								
Turnover		1	0.076	0.787		1	0.007	0.934
C. carassius	0.529 ± 0.006				0.363 ± 0.012			
No *C. carassius*	0.538 ± 0.003				0.367 ± 0.005			
Nestedness		1	5.844	**0.028**		1	0.595	0.452
C. carassius	0.070 ± 0.001				0.100 ± 0.008			
No *C. carassius*	0.033 ± 0.001				0.074 ± 0.003			
Total beta diversity		1	0.095	0.762		1	0.020	0.890
C. carassius	0.572 ± 0.003				0.432 ± 0.005			
No *C. carassius*	0.565 ± 0.001				0.436 ± 0.001			
eDNA metabarcoding								
Turnover		1	2.295	0.159		1	0.682	0.421
C. carassius	0.505 ± 0.013				0.364 ± 0.012			
No *C. carassius*	0.440 ± 0.003				0.328 ± 0.005			
Nestedness		1	0.646	0.434		1	0.996	0.333
C. carassius	0.083 ± 0.002				0.115 ± 0.010			
No *C. carassius*	0.067 ± 0.001				0.075 ± 0.005			

Total beta diversity		1	5.853	**0.028**		1	4.238	0.056
C. carassius	0.570 ± 0.007				0.459 ± 0.005			
No *C. carassius*	0.492 ± 0.003				0.402 ± 0.002			
Combined methods								
Turnover		1	3.046	0.100		1	1.223	0.285
C. carassius	0.518 ± 0.003				0.360 ± 0.010			
No *C. carassius*	0.470 ± 0.004				0.317 ± 0.003			
Nestedness		1	0.291	0.597		1	0.129	0.724
C. carassius	0.033 ± 0.001				0.068 ± 0.003			
No *C. carassius*	0.041 ± 0.001				0.059 ± 0.002			
Total beta diversity		1	5.617	**0.031**		1	7.854	**0.013**
C. carassius	0.549 ± 0.002				0.431 ± 0.003			
No *C. carassius*	0.507 ± 0.001				0.374 ± 0.001			
Method comparison								
Turnover		2	2.340	0.107		2	0.260	0.772
Microscopy	0.559 ± 0.005				0.379 ± 0.010			
DNA metabarcoding	0.556 ± 0.003				0.385 ± 0.009			
eDNA metabarcoding	0.511 ± 0.009				0.362 ± 0.010			
Nestedness		2	0.731	0.486		2	0.889	0.417
Microscopy	0.047 ± 0.001				0.113 ± 0.005			
DNA metabarcoding	0.048 ± 0.001				0.081 ± 0.004			
eDNA metabarcoding	0.060 ± 0.002				0.091 ± 0.007			
Total beta diversity		2	1.659	0.200		2	0.242	0.786
Microscopy	0.595 ± 0.003				0.466 ± 0.007			
DNA metabarcoding	0.589 ± 0.001				0.455 ± 0.002			
eDNA metabarcoding	0.561 ± 0.006				0.450 ± 0.005			

At species-level, sweep-netting with microscopy and eDNA metabarcoding revealed a weak positive effect of *C. carassius* presence on turnover (Figs. 5.3ai, iii) and total beta diversity (Figs. 5.3ci, iii) between ponds, but not nestedness (Figs. 5.3bi, iii). In contrast, DNA metabarcoding did not identify a significant effect of *C. carassius* presence on turnover, nestedness, or total beta diversity (Figs. 5.3aii, bii, cii). At family-level, DNA metabarcoding and eDNA metabarcoding revealed a weak or moderate positive influence of *C. carassius* presence on turnover (Figs. 5.4aii, iii) and total beta diversity (Figs. 5.4cii, iii), but not nestedness (Figs. 5.4bii, iii). Yet, no significant effect of *C. carassius* presence

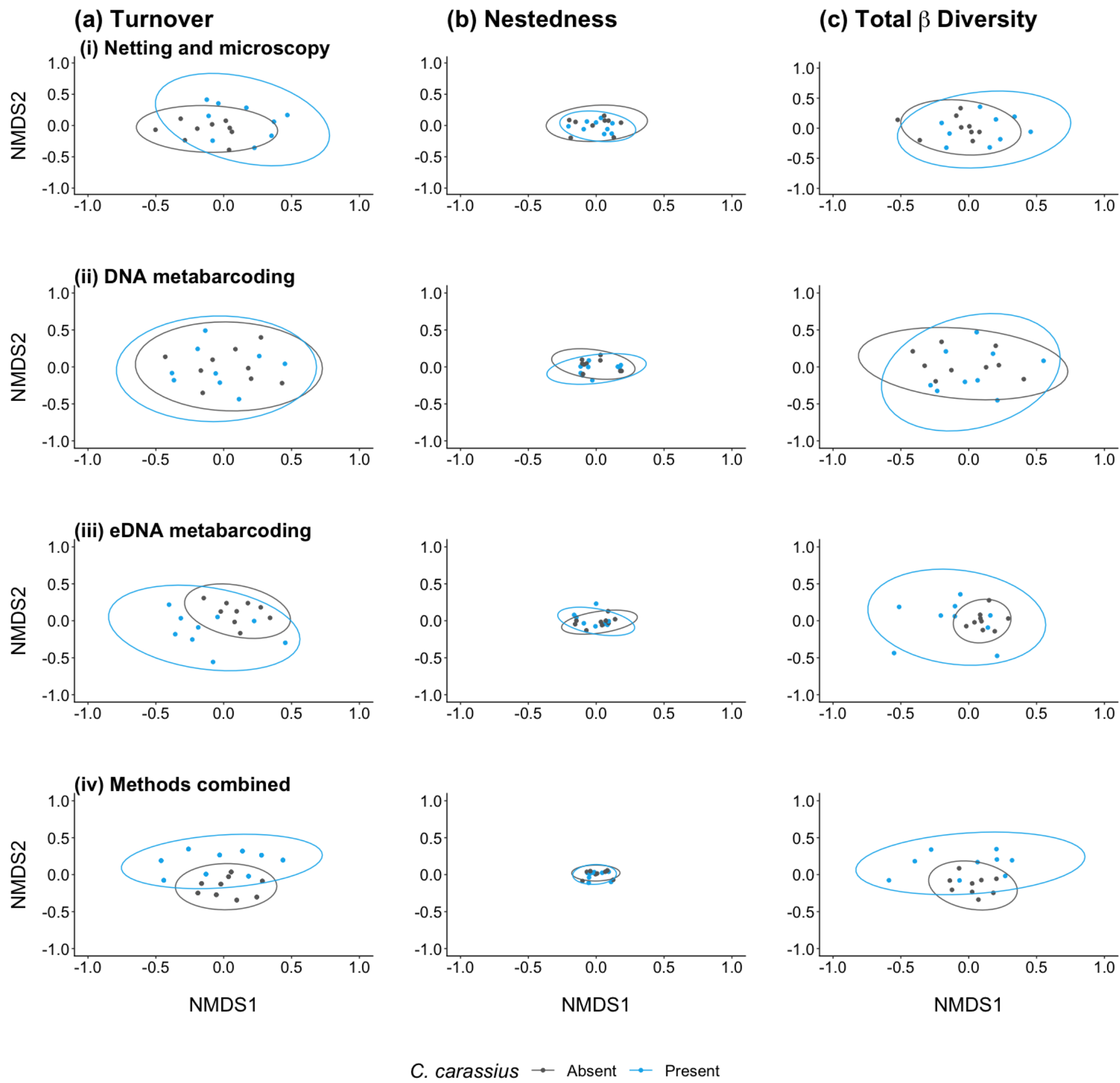


Figure 5.3: Non-metric Multidimensional Scaling (NMDS) plots of species-level invertebrate communities (Jaccard dissimilarity) from ponds with *C. carassius* (blue points/ellipse) and without fish (grey points/ellipse) across Norfolk and East Yorkshire. The turnover **(a)** and nestedness **(b)** partitions of total beta diversity **(c)** are shown according to method of invertebrate assessment: netting and microscopy **(i)**, DNA metabarcoding **(ii)**, eDNA metabarcoding **(iii)**, and all methods combined **(iv)**.

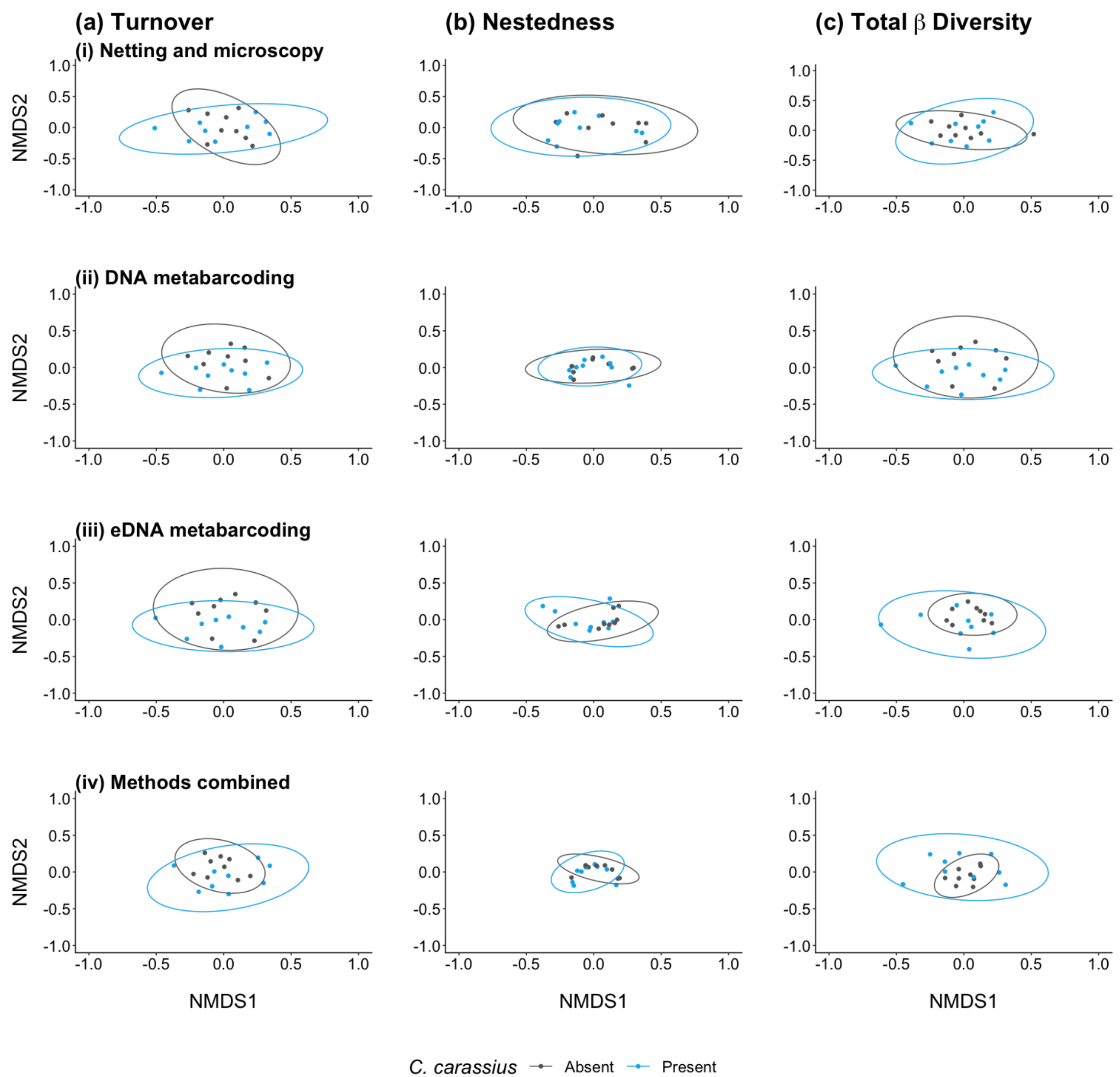


Figure 5.4: Non-metric Multidimensional Scaling (NMDS) plots of family-level invertebrate communities (Jaccard dissimilarity) from ponds with *C. carassius* (blue points/ellipse) and without fish (grey points/ellipse) across Norfolk and East Yorkshire. The turnover **(a)** and nestedness **(b)** partitions of total beta diversity **(c)** are shown according to method of invertebrate assessment: netting and microscopy **(i)**, DNA metabarcoding **(ii)**, eDNA metabarcoding **(iii)**, and all methods combined **(iv)**.

on turnover, nestedness, or total beta diversity was found using sweep-netting and microscopy (Figs. 5.4ai, bi, ci). Broadly, sweep-netting/microscopy and eDNA metabarcoding produced concurrent results at species-level, whereas DNA metabarcoding and eDNA metabarcoding were more concordant at family-level. However, congruence between methods changed depending on the partition of beta

diversity being investigated. All methods combined revealed *C. carassius* presence had moderate and strong positive effects on turnover and total beta diversity at species-level (Figs. 5.3aiv, civ) and family-level (Figs. 5.4aiv, civ) respectively, but not on nestedness (Table 5.4).

Table 5.4: Summary of analyses (PERMANOVA) statistically examining variation in community composition of ponds with and without *C. carassius*, and across methods at species-level and family-level.

	Community similarity (PERMANOVA)							
	Species-level				Family-level			
	df	***F***	***R²***	***P***	**df**	***F***	***R²***	***P***
Netting and microscopy								
Turnover	1	1.673	0.095	**0.030**	1	1.198	0.070	0.370
Nestedness	1	-3.454	-0.275	0.961	1	-0.103	-0.007	0.631
Total beta diversity	1	1.369	0.079	**0.039**	1	1.136	0.066	0.307
DNA metabarcoding								
Turnover	1	1.304	0.075	0.126	1	2.038	0.113	**0.020**
Nestedness	1	-2.666	-0.200	0.951	1	-1.136	-0.076	0.906
Total beta diversity	1	1.134	0.066	0.210	1	1.528	0.087	**0.049**
eDNA metabarcoding								
Turnover	1	2.484	0.134	**0.002**	1	1.850	0.104	**0.021**
Nestedness	1	-2.136	-0.154	0.946	1	-0.015	-0.001	0.708
Total beta diversity	1	1.841	0.103	**0.002**	1	1.521	0.087	**0.032**
Combined methods								
Turnover	1	1.958	0.109	**0.001**	1	1.777	0.100	**0.020**
Nestedness	1	-1.687	-0.118	0.955	1	0.966	0.057	0.417
Total beta diversity	1	1.683	0.095	**0.001**	1	1.567	0.089	**0.013**
Method comparison								
Turnover	2	6.721	0.209	**0.001**	2	15.936	0.385	**0.001**
Nestedness	2	-9.762	-0.620	1.000	2	-6.738	-0.359	1.000
Total beta diversity	2	5.057	0.166	**0.001**	2	10.808	0.298	**0.001**

Additional analyses undertaken on data from the sampling methods combined supported an effect of *C. carassius* presence-absence and excluded the influence of abiotic variables on pond invertebrate diversity. At species-level, only pond area was identified as a

significant abiotic variable for turnover and total beta diversity by forward selection. No significant abiotic variables were identified for nestedness. Consequently, variance partitioning analysis was only undertaken for turnover and total beta diversity using *C. carassius* presence-absence and pond area. Based on the adjusted R^2 values, biotic and abiotic variables explained 6.45% and 4.35% of the total variation in turnover and total beta diversity respectively (Fig. 5.5). *C. carassius* presence-absence made a significant contribution to turnover (Fig. 5.5a; adjusted R^2 = 5.34%, F_1 = 1.656, P = 0.008) and total beta diversity (Fig. 5.5b; adjusted R^2 = 3.86%, F_1 = 1.438, P = 0.011), whereas pond area explained less variance and did not significantly impact species turnover (Fig. 5.5a; adjusted R^2 = 2.61%, F_1 = 1.191, P = 0.175) or total beta diversity (Fig. 5.5b; adjusted R^2 = 1.73%, F_1 = 1.082, P = 0.271). RDA of nestedness without abiotic data indicated no impact of *C. carassius* presence-absence (F_1 = 0.3244, P = 0.877). At family-level, forward selection did not identify any significant abiotic variables for turnover, nestedness, or total beta diversity. Therefore, variance partitioning was not undertaken for any component of beta diversity. RDA of each beta diversity component minus abiotic data revealed *C. carassius* presence-absence influenced turnover (F_1 = 1.633, P = 0.025) and total beta diversity (F_1 = 1.567, P = 0.015), but not nestedness (F_1 = 0.956, P = 0.376).

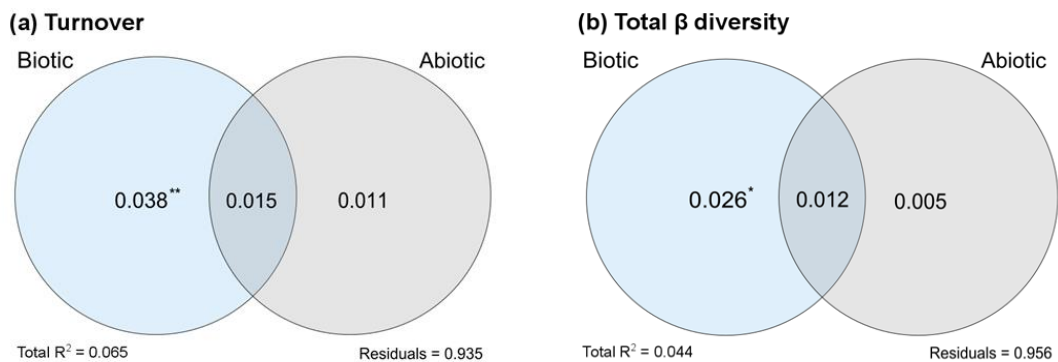


Figure 5.5: The relative contribution of biotic (*C. carassius* presence-absence) and abiotic variables (pond area) to species-level turnover (a) and total beta diversity (b) when the combined invertebrate data from all three sampling methods were considered. Values within circles and circle intersections represent the adjusted R^2 values. Significant variables are indicated by asterisks (* = $P < 0.05$, ** = $P < 0.01$). Negative fraction values are not presented.

5.3.3 Comparison of methods for freshwater invertebrate assessment

Only 17 species (Fig. 5.6a) and 22 families (Fig. 5.6b) were detected by all three methods of invertebrate assessment. eDNA metabarcoding detected the most unique species and families, whereas DNA metabarcoding and morphotaxonomic identification were more comparable. There was no overlap between sweep-netting with microscopy and eDNA metabarcoding at either taxonomic rank as opposed to DNA and eDNA metabarcoding, which shared 50 species and 18 families. Sweep-netting with microscopy and DNA metabarcoding were similar at family-level with an overlap of 13 families, but dissimilar at species-level, with a roughly equal number of shared and unique species records.

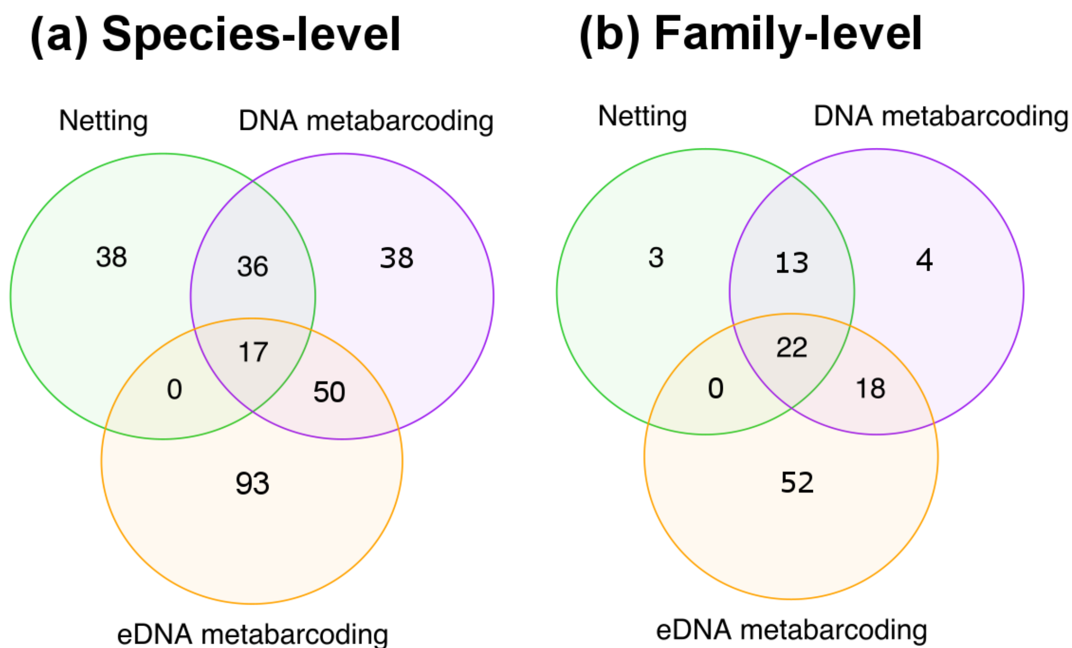


Figure 5.6: Venn diagram which summarises the number of invertebrate species (a) and families (b) detected across the 18 study ponds by each method of invertebrate assessment: sweep-netting and microscopy (green circle), DNA metabarcoding (purple circle), and eDNA metabarcoding (orange circle). Overlap in species or family detections between methods is displayed within circle intersections.

Sampling method had a significant effect on alpha diversity of ponds at species-level (GLM χ^2_2 = 36.243, P < 0.001) and family-level (GLM χ^2_2 = 54.658, P < 0.001).

Significant differences between the alpha diversity means of sweep-netting with microscopy and DNA metabarcoding (species-level -0.467 ± 0.132, Z = -3.534, P = 0.001; family-level -0.302 ± 0.108, Z = -2.791, P = 0.015), sweep-netting with microscopy and eDNA metabarcoding (species-level -0.779 ± 0.130, Z = -6.009, P < 0.001; family-level -0.729 ± 0.102, Z = -7.128, P < 0.001), and DNA and eDNA metabarcoding (species-level 0.312 ± 0.124, Z = 2.521, P = 0.031; family-level 0.427 ± 0.096, Z = 4.454, P < 0.001) were observed. Alpha diversity was lower using sweep-netting and microscopy than either metabarcoding approach, and higher using eDNA metabarcoding than DNA metabarcoding (Fig. 5.7). MVDISP was not significantly different between methods for turnover, nestedness, or total beta diversity at either taxonomic rank (Table 5.2). Sampling method had moderate and strong positive effects on turnover and total beta diversity at species-level (Figs. 5.8ai, ci) and family-level (Figs. 5.8aii, cii) respectively, but not nestedness (Table 5.4).

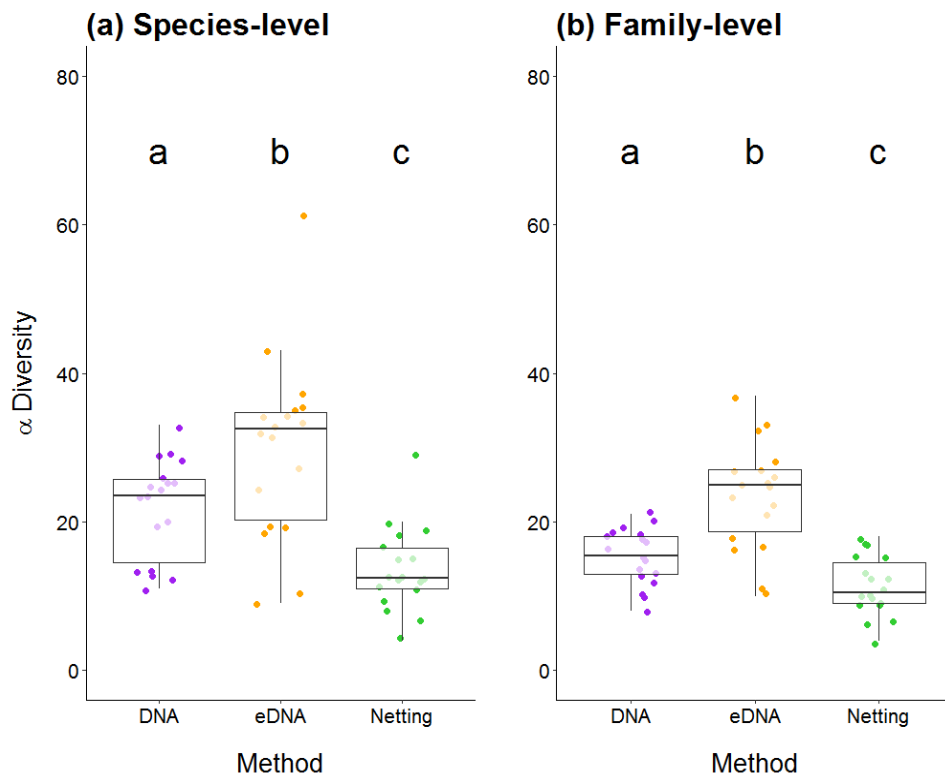


Figure 5.7: Mean alpha diversity (taxon richness) of invertebrates in ponds across Norfolk and East Yorkshire. Alpha diversity at species-level **(a)** and family-level **(b)** is displayed according to method of invertebrate assessment: netting and microscopy (green points), DNA metabarcoding (purple points), and eDNA metabarcoding (orange points). Boxes show 25th, 50th, and 75th percentiles, and whiskers show 5th and 95th percentiles.

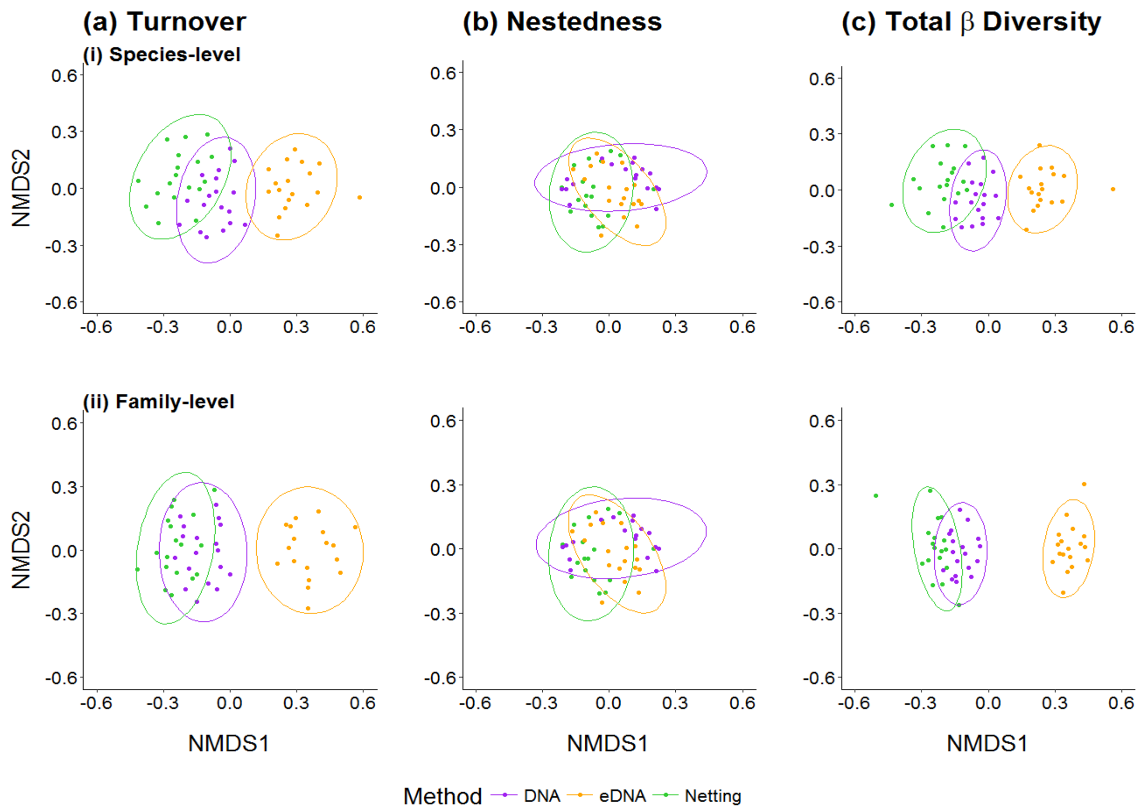


Figure 5.8: Non-metric Multidimensional Scaling (NMDS) plots of invertebrate communities (Jaccard dissimilarity) produced by sweep-netting and microscopy (green points/ellipse), DNA metabarcoding (purple points/ellipse), and eDNA metabarcoding (orange points/ellipse) for the 18 study ponds. The turnover **(a)** and nestedness **(b)** partitions of total beta diversity **(c)** are shown at species-level **(i)** and family-level **(ii)**.

5.4 Discussion

We have demonstrated that *C. carassius* has a negligible impact on invertebrate taxon richness of ponds, but may enhance beta diversity of pond networks by inducing turnover in invertebrate community composition. Our results corroborate previous work on *C. carassius* (Stefanoudis *et al.*, 2017) and imply that stocking of this species for conservation management (Copp & Sayer, 2010) should continue due to the invertebrate species and families exclusive to ponds with or without *C. carassius* respectively. We also found that three different methods of pond invertebrate assessment were complementary, and yielded different species and family inventories. Consequently, these methods should be used in combination to provide the most complete picture of

invertebrate diversity and best inform freshwater management.

5.4.1 Impact of *C. carassius* stocking on pond invertebrates

C. carassius had a negligible influence on alpha diversity and positive influence on beta diversity of ponds in terms of invertebrates. Total alpha diversity in ponds with *C. carassius* was marginally reduced compared to total alpha diversity in fishless ponds, but this difference was not significant across methods used at either taxonomic rank. Within the major invertebrate groups identified by all methods combined, species-level alpha diversity of Coleoptera and Mollusca was significantly reduced in ponds containing *C. carassius* as opposed to fishless ponds. Across ponds, total beta diversity of invertebrate communities was driven by turnover (taxon substitution) rather than nestedness (taxon subsets) (Baselga & Orme, 2012). Detailed analyses revealed *C. carassius* presence-absence positively influenced turnover and total beta diversity between ponds. Therefore, taxa in fishless ponds were replaced by different taxa in ponds with *C. carassius*, resulting in dissimilar community composition.

Our results both echo and contradict those of Stefanoudis *et al*. (2017), where the presence of fish (including *C. carassius*) in ponds altered macrophyte and cladoceran community composition, but not water beetle composition. Hassall *et al*. (2011) also found that fish presence in ponds had a positive effect on species richness of most invertebrate taxa, excluding a negative effect on Coleoptera species richness. This reaffirms results from Gee *et al*. (1997) who observed no influence of fish stocking on macrophyte and macroinvertebrate species richness, albeit Odonata richness was lower and Trichoptera richness higher in stocked ponds. Conversely, other research found that managed/stocked ponds, some of which contained *C. carassius*, had lower invertebrate diversity than unmanaged sites, which were characterised by Trichoptera, Coleoptera and Zygoptera larvae (Wood *et al.*, 2001). Similarly, large, active and free-swimming taxa (Notonectidae, Corixidae, Gyrinidae, Dytiscidae, Aeshnidae, Libellulidae and Chaoboridae) were strongly associated with fish absence as well as more diverse and abundant in fishless lakes (Bendell & McNicol, 1995; Schilling *et al.*, 2009a, b). Here, we found higher Coleopteran diversity was associated with *C. carassius* absence.

Critically, few of the aforementioned studies accounted for the identity of fish species assemblages present in ponds (Wood *et al.*, 2001; Stefanoudis *et al.*, 2017), whereas other studies only accounted for the presence of a particular species (Schilling *et*

al., 2009a) or fish presence-absence generally (Bendell & McNicol, 1995; Gee *et al.*, 1997; Schilling *et al.*, 2009b; Hassall *et al.*, 2011). The contrasting results produced by these studies and our own would indicate that the impact of fish stocking on pond biodiversity is highly dependent on the species stocked and management strategy. Wetland fishes vary in dietary preference and consume different proportions of invertebrate taxa, thus different fish species will suppress numbers of and confer benefits to different invertebrate taxa (Batzer, Pusateri & Vetter, 2000). Invasive species may be more detrimental than non-invasive species, for example, the mosquitofish (*Gambusia affini*) reduced zooplankton abundance and macroinvertebrate density by 90% and 50% respectively after introduction in a wetland ecosystem experiment (Preston *et al.*, 2017). Regarding management strategy, the duration of stocking was found to substantially reduce invertebrate species richness and abundance (Schilling *et al.*, 2009a). Therefore, local- and regional-scale diversity may benefit most from ponds that are regularly drained and fish-free, or ponds that are regularly drained and stocked with fish at low biomass (Lemmens *et al.*, 2013).

In addition to *C. carassius* presence-absence, the effects of environmental variables on alpha and beta diversity of pond invertebrates must be considered (Hassall *et al.*, 2011). The ponds in our study were selected to be similar in their physical and chemical properties, which may explain the lack of or minimal contribution of abiotic factors to variance in invertebrate community structure. Although pond area was retained by model selection, this variable did not significantly influence community structure and explained less variance than *C. carassius* presence-absence. Other studies have also shown a weak or no effect of pond area on invertebrate species richness and community composition (Gee *et al.*, 1997; Wood *et al.*, 2001; Oertli *et al.*, 2002; Gledhill *et al.*, 2008). Critically, our environmental data were collected over a 7-year period, whereas contemporary data may have explained more variance in community structure. We did not include variables that experience high temporal variation (e.g. temperature, pH, nutrient concentration, and surface dissolved oxygen) in our analyses, but these may have contributed to differences in community structure. Large, free-swimming invertebrates that are vulnerable to fish predation may be more abundant in ponds with acidic conditions that fish cannot tolerate (Bendell & McNicol, 1995). Invertebrate species richness, particularly Coleoptera and Gastropoda, was negatively correlated with nutrient concentration (i.e. eutrophication) of ponds (Menetrey *et al.*, 2005; Hassall *et al.*, 2011). Invertebrate communities are also sensitive to oxygen depletion, a side effect of nutrient

enrichment (Menetrey *et al.*, 2005), and invertebrate richness was found to be lower at intermediate levels of oxygen demand (Hassall *et al.*, 2011). In contrast, *C. carassius* can tolerate anoxic conditions (Sayer *et al.*, 2011; Stefanoudis *et al.*, 2017). The alpha diversity reductions for Coleoptera and Mollusca in ponds with *C. carassius* may be linked to the aforementioned variables. Therefore, investigations examining contemporary physicochemical variables in combination with *C. carassius* presence-absence would be highly valuable to disentangle the impact of stocking from habitat associations.

C. carassius is often assumed to have negative impacts on pond biodiversity like other cyprinids. Foraging activity of the common carp (*Cyprinus carpio*) especially reduces invertebrate density and macrophyte cover, which has knock-on effects for waterfowl species richness and abundance (Haas *et al.*, 2007; Maceda-Veiga *et al.*, 2017). Similarly, the diversity and richness of invertebrates and macrophytes dictates amphibian foraging and reproductive success (Rannap & Briggs, 2006; Gustafson *et al*., 2006). Our study supports *C. carassius* as a potentially important and positive driver of community heterogeneity and subsequently beta diversity in ponds. Our findings indicate that stocking of *C. carassius* may enhance invertebrate diversity across pond networks, and that current management of stocked populations is appropriate. However, the impact of *C. carassius* on pond biodiversity must be studied more broadly with respect to *C. carassius* population density and environmental variables (e.g. water temperature, pH, surface dissolved oxygen). Effects of this fish species on amphibians, waterfowl, and mammals as well as invertebrates utilising ponds must be assessed to determine whether stocking is truly beneficial.

5.4.2 Comparison of methods for freshwater invertebrate assessment

eDNA metabarcoding generated the highest alpha diversity at species-level and family-level, followed by DNA metabarcoding, then sweep-netting and microscopy. However, each method of invertebrate assessment detected unique species and produced a different community. Sweep-netting with microscopy was more similar in community composition to DNA metabarcoding, and these methods performed best for Coleoptera, Hirudinea, Megaloptera, and Odonata. In contrast, eDNA metabarcoding produced a markedly different community and detected taxa that are typically overlooked in or missed entirely from netted samples, including Arachnida, Cnidaria, Crustacea, Hymenoptera,

Lepidoptera, Platyhelminthes, and Rotifera. Nonetheless, DNA and eDNA metabarcoding were comparable in terms of performance for Annelida, Collembola, and Diptera, particularly at species-level. Despite failure to recover some taxa, our results reinforce other studies where metabarcoding captured more diversity than conventional morphotaxonomic approaches and resolved problematic groups (e.g. Diptera) that are difficult to morphologically identify to species (Elbrecht *et al.*, 2017b; Clarke *et al.*, 2017; Klymus *et al.*, 2017a; Emilson *et al.*, 2017; Andújar *et al.*, 2017; Lobo *et al.*, 2017; Carew *et al.*, 2018b).

The differences between these methods are not wholly unexpected due to their inherent biases. The UK National Pond Survey methodology recommends that netted samples are placed into buckets, followed by sorting and identification in the laboratory as soon as possible. Samples can be refrigerated to prolong processing time, but must still be processed within three days of collection (Biggs *et al.*, 1998). In our study, all ponds were netted in a single day and would not have been processed in the recommended time frame. This standard methodology was also not designed with the use of specimens for molecular applications in mind. After collection, we immediately placed samples on ice for transport to the laboratory, whereupon they were frozen at -20 °C. This to prevent predation within samples during transport, minimise organismal decay and subsequent DNA degradation, and allow samples to be processed as and when required in the laboratory. Our strategy meant that small or inconspicuous dead specimens would not have been recovered during sorting (Biggs *et al.*, 1998) and thus excluded from the microscopy and DNA metabarcoding inventories. These losses are surplus to the 29% specimens typically overlooked during sorting due to smaller body size (Haase *et al.*, 2010). Some recovered specimens may also have been damaged or completely destroyed by sorting, influencing morphotaxonomic identification (Lobo *et al.*, 2017; Zizka *et al.*, 2018).

The pre-sorting treatment and sorting process likely contributed to the differences between sweep-netting and microscopy, DNA metabarcoding, and eDNA metabarcoding. Another source of discrepancy between species-level sweep-netting with microscopy and metabarcoding (particularly DNA) is human error during identification. Taxa may have been omitted from species-level inventories as they can only be reliably identified to genus or family-level, or taxa may have been falsely identified and inventoried (Haase *et al.*, 2010; Carew *et al.*, 2013; Elbrecht *et al.*, 2017b). The losses incurred by the pre-

sorting treatment and sorting process for morphotaxonomic identification cannot be mitigated, and human error can only be reduced with taxonomic expertise. Conversely, species recovery by DNA metabarcoding can be improved through the development of protocols that preserve and extract DNA from intact, unsorted samples, thereby removing the time consuming, precarious sorting step (Elbrecht *et al.*, 2017c). For example, the ethanol used to preserve specimens (Zizka *et al.*, 2018), or temporary replacement of ethanol with DNA extraction buffer in specimen tubes followed by incubation then specimen removal (Carew, Coleman & Hoffmann, 2018a), offer alternative starting materials to bulk tissue for DNA metabarcoding. However, these alternative sources of DNA tend to produce false negatives for schlerotised groups, such as Coleoptera and Trichoptera (Zizka *et al.*, 2018; Carew *et al.*, 2018a).

Metabarcoding is subject to a number of other biases that stem from different stages of the workflow, predominantly DNA quality, marker choice, primer design, amplification bias, species masking, and reference databases (Taberlet *et al.*, 2012; Deiner *et al.*, 2017). Although the DNA metabarcoding approach has been successfully applied to bulk tissue samples for dietary analysis (Trevelline *et al.*, 2018) and biomonitoring (Elbrecht *et al.*, 2017b; Emilson *et al.*, 2017; Andújar *et al.*, 2017; Macher *et al.*, 2018; Carew *et al.*, 2018b), recurring issues have been encountered. Arguably, the most pressing issue is that of size bias, where DNA from large and/or high biomass taxa can outcompete DNA of smaller and/or low biomass taxa during PCR amplification and sequencing (Elbrecht *et al.*, 2017a). To minimise this bias, we used the size sorting approach conceived by Elbrecht *et al.* (2017a) and sequenced body size categories independently, followed by data pooling downstream. Despite these countermeasures, DNA metabarcoding failed to detect some taxa that have reference sequences and can be reliably identified by microscopy, including several Coleopterans (*Agabus sturmii*, *Hydroporus erythrocephalu*s, *Rhantus frontalis*, *Haliplus confinis*, *Haliplus ruficollis*), a small Mollusc (*Gyraulus crista*), a medium Hirudinean (*Erpobdella lineata*), and two large Anisopterans (*Aeshna mixta*, *Anax imperator*); although, different species of *Erpobdella* and *Aeshna* amplified when tested *in vitro*. The non-delivery of desired results by size-sorting is problematic as this process is time-consuming, labour-intensive, and potentially increases cross-contamination risk between samples (Elbrecht *et al.*, 2017a). It also does not eliminate the cumbersome sorting of samples from vegetation and substrate (Elbrecht *et al.*, 2017b). Size-sorting may therefore be a drain on resources and time allocated to DNA metabarcoding projects, but other sources of false negatives must

be excluded before size-sorting is deemed redundant.

In contrast to DNA metabarcoding, eDNA metabarcoding of water preferentially amplifies DNA from planktonic organisms (Deiner *et al.*, 2016; Macher *et al.*, 2018), and these organisms are often retained on filter membranes used for eDNA capture (*pers. obs.*). Consequently, this DNA is abundant in samples and overwhelms DNA from other taxa during sequencing. Pre-filtering with a large pore size filter membrane or only using large pore size filter membranes may reduce this particular bias (Macher *et al.*, 2018). Alternatively, different sources of eDNA should be considered. Sediment has been used in eDNA metabarcoding assessments of marine invertebrates (Aylagas *et al.*, 2018), but has not been applied in freshwater metabarcoding studies. Given the diversity of benthic invertebrates present in these ecosystems, eDNA may be more abundunt in sediment than the water column (Klymus *et al.*, 2017a). Therefore, future assessments of freshwater invertebrates using metabarcoding should be made based on sediment and water samples.

Another possible explanation for the different taxa detected by eDNA metabarcoding is variability in eDNA production and shedding rates across species. Indeed, the species that were infrequently detected by eDNA metabarcoding were those that possess thicker exoskeletons composed of chitin and occasionally calcium carbonate, e.g. Coleoptera, Hemiptera. Exoskeletons may restrict the release of DNA into the water column (Tréguier *et al.*, 2014) as opposed to organisms that are filter-feeders or produce slime, such as Crustacea and Mollusca, ectoparasites that feed on blood or skin of other species (i.e. Acari), or use external instead of internal fertilisation. Different species also have different habitat preferences within freshwater ecosystems and their utilisation of these habitats may vary, potentially resulting in a highly localised distribution (Klymus *et al.*, 2017a). Therefore, more samples or greater volumes may be required to improve detection probability of pond biota (Harper *et al.*, 2019a).

Metabarcoding marker choice and primer design can substantially influence amplification success and taxonomic assignment. Although the *COI* region offers species resolution, has extensive database representation, and is used as standard in DNA barcoding (Elbrecht, Hebert & Steinke, 2018), it lacks conserved primer-binding sites as a protein-coding gene (Clarke *et al.*, 2017). This is problematic for metabarcoding of diverse species assemblages due to high risk of primer mismatch and subsequent bias (Clarke *et al.*, 2017; Elbrecht *et al.*, 2018). Primer bias may prevent the recovery of all taxa present due to preferential amplification of DNA from particular taxa (Elbrecht *et al.*, 2016; Lobo *et al.*, 2017). Consequently, most metabarcoding primers designed to

target the *COI* region in metazoans (Meusnier *et al.*, 2008; Zeale *et al.*, 2011; Geller *et al.*, 2013; Leray *et al.*, 2013; Elbrecht & Leese, 2017) are degenerate to allow primers to bind at highly variable sites. However, high degeneracy may allow primers to bind non-target regions (Elbrecht *et al.*, 2018) and create biased amplification toward non-metazoans, e.g. bacteria, algae, fungi (Brandon-Mong *et al.*, 2015; Macher *et al.*, 2018). This bias can occur even when primers are designed to amplify a specific metazoan group, such as invertebrates (Elbrecht & Leese, 2017; Elbrecht *et al.*, 2017b), and is more pronounced in sequences obtained from eDNA samples (Macher *et al.*, 2018). Macher *et al*. (2018) suggest that non-target amplification is beneficial for identifying new bioindicators, but unintended amplification can induce false negatives and lead to misinformation in freshwater management.

Other strategies have been put forward to mitigate amplification bias in metabarcoding studies, such as optimisation of thermocycling conditions (Clarke *et al.*, 2017). Clarke *et al*. (2017) found that a consistent annealing temperature (46 °C) and reduced PCR cycle number substantially improved zooplankton species detection by primers mlCOIintF and jgHCO2198 as opposed to the touchdown PCR conditions set out by Leray *et al*. (2013). Here, we also used a consistent annealing temperature (47 °C) and reduced cycle number for the aforementioned primers, thus thermocycling protocol is an unlikely source of amplification bias in our study. Alternatively, metabarcoding performance for invertebrates may be improved with the use of different or multiple markers (Deiner *et al.*, 2016). Indeed, Elbrecht *et al*. (2016) found amplification bias was reduced with 16S ribosomal rRNA as opposed to *COI* for freshwater invertebrate bulk tissue samples. Unfortunately, these alternative markers often provide less taxonomic resolution due to lack of reference database representation (Elbrecht *et al.*, 2016; Clarke *et al.*, 2017). Use of multiple markers will also inevitably increase PCR and sequencing costs, restricting the application of metabarcoding to freshwater monitoring schemes (e.g. Environment Agency, Freshwater Habitats Trust, Riverfly Partnership, British Dragonfly Society), but sequencing costs are expected to subside in the future (Elbrecht *et al.*, 2017b).

Similar to Macher *et al*. (2018), we obtained an excessive number of unassigned reads from eDNA metabarcoding (almost 60%), despite using a relaxed BLAST identity (90%) against our custom database and the entire NCBI nucleotide database. This suggests that sequences were of poor quality, could not be assigned to any reference sequences, or lacked reference database representation (Macher *et al.*, 2018). Recent

research by Elbrecht *et al*. (2018) demonstrated that degenerate primers (such as mlCOIintF designed by Leray *et al*., 2013 and used here) can experience primer slippage and produce sequences of variable length when they bind to DNA regions containing low diversity. This is variable across species and has implications for bioinformatic processing and eventual taxonomic assignment. The reliance on public reference databases is also problematic for taxonomic assignment, even for species with sequence records. Public records may be few, mislabelled, or have limited geographic coverage (Elbrecht *et al.*, 2017b; Klymus *et al.*, 2017a; Curry *et al.*, 2018). In our study, we identified 19 species belonging to Hemiptera, Mollusca, and Odonata using microscopy that were not represented in reference sequence databases. DNA from these species may have been amplified and sequenced by metabarcoding, but sequences would not have been taxonomically assigned to species-level. Researchers and practitioners must focus on the development of more specific primers that have binding and flanking regions of high nucleotide diversity (Elbrecht *et al.*, 2018) and/or target particular invertebrate orders or families (Klymus *et al.*, 2017a) as well as the procurement of reference sequences for different markers (Elbrecht *et al.*, 2016; Curry *et al.*, 2018). These are essential steps to improve the reliability and accuracy of molecular monitoring in freshwater ecosystems.

5.4.3 Concluding remarks

Using a multi-method approach, we have demonstrated that *C. carassius* has a different impact on invertebrate diversity to other fishes typically stocked in ponds. This has implications for the conservation of *C. carassius* and pond ecosystems. Fish are generally perceived to negatively impact pond biodiversity, particularly invertebrates and amphibians. Yet, *C. carassius* appears to have a negligible influence on invertebrate diversity in individual ponds, and could benefit invertebrate diversity across pond networks by introducing community heterogeneity. This would imply that stocking of *C. carassius* should continue to conserve this species and pond biodiversity. However, there is a need to evaluate the impact of *C. carassius* stocking on other pond biota, including amphibians, waterfowl, and mammals. Our findings also highlight the potential of molecular tools for freshwater invertebrate assessment and ecological investigation. Importantly, sweep-netting and microscopy, DNA metabarcoding, and eDNA metabarcoding all revealed unique invertebrate diversity present in ponds. Therefore,

these tools should be used in combination for freshwater monitoring and research to reliably inform conservation and management decisions.

5.5 Acknowledgements

This work was funded by the University of Hull. We would like to thank Michael Lee and Matt Smith (Environment Agency) for selecting East Yorkshire ponds and providing associated fyke net data. We would also like to express our gratitude to Jane Colley (Askham Bryan College) for providing invertebrate samples from these ponds. We are grateful to Jianlong Li and Cristina Di Muri (University of Hull) for assistance with water sampling and filtration as well as advice on library preparation and sequencing.

5.6 Data accessibility

Raw sequence reads have been archived on the NCBI Sequence Read Archive (Study: SRP163672; BioProject: PRJNA494857; BioSamples: SAMN10181701 - SAMN10182084 [bulk tissue DNA] and SAMN10187732 - SAMN10188115 [eDNA]; SRA accessions: SRR7969394 - SRR796977 [bulk tissue DNA] and SRR7985814 - SRR7986197 [eDNA]). Jupyter notebooks, R scripts and corresponding data are deposited in a dedicated GitHub repository (https://github.com/lrharper1/LRHarper_PhDThesis_Chapter5) which has been permanently archived (https://doi.org/10.5281/zenodo.2634240).

Chapter 6: Environmental DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority mammals

European otters (*Lutra lutra*) (Linnaeus, 1758) emerging from pond

© user: Peter Trimming| Flickr | CC BY 2.0

This chapter is available online as

Harper, L.R., Handley, L.L., Carpenter, A.I., Ghazali, M., Di Muri, C., Macgregor, C.J., Logan, T.W., Law, A., Breithaupt, T., Read, D.S., McDevitt, A.D. & Hänfling, B. (2019) Environmental DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority mammals. *bioRxiv*, 546218. https://doi.org/10.1101/546218

Abstract

Environmental DNA (eDNA) metabarcoding is largely used to survey aquatic communities, but can also provide data on terrestrial taxa utilising aquatic habitats. However, the entry, dispersal, and detection of terrestrial species' DNA within water bodies is understudied. We evaluated eDNA metabarcoding of pond water for monitoring semi-aquatic, ground-dwelling, and arboreal mammals, and examined spatiotemporal variation in mammal eDNA signals using experiments in captive and wild conditions. We selected nine focal species of conservation and management concern: European water vole (*Arvicola amphibius*), European otter (*Lutra lutra*), Eurasian beaver (*Castor fiber*), European hedgehog (*Erinaceus europaeus*), European badger (*Meles meles*), red deer (*Cervus elaphus*), Eurasian lynx (*Lynx lynx*), red squirrel (*Sciurus vulgaris*), and European pine marten (*Martes martes*). We hypothesised that eDNA signals (i.e. proportional read counts) would be stronger for semi-aquatic than terrestrial species, and at sites where mammals exhibited behaviours (e.g. swimming, urination). We tested this by sampling water bodies in captive focal species enclosures at specific sites where behaviours had been observed ('directed' sampling) and at equidistant intervals along the shoreline ('stratified' sampling). We surveyed natural ponds ($N = 6$) where focal species were present using stratified water sampling, camera traps, and field signs. eDNA samples were metabarcoded using vertebrate-specific primers. All focal species were detected in captivity. eDNA signal strength did not differ between directed and stratified samples across or within species, between species lifestyles (i.e. semi-aquatic, ground-dwelling, arboreal), or according to behaviours. Therefore, eDNA was evenly distributed within artificial waterbodies. Conversely, eDNA was unevenly distributed in natural ponds. eDNA metabarcoding, camera trapping, and field signs shared three species detections, but eDNA metabarcoding missed two species were recorded with cameras and field signs. Nonetheless, eDNA metabarcoding detected small mammals missed by cameras and field signs, e.g. *A. amphibius*. Terrestrial mammal eDNA signals were weaker and detected in fewer samples than semi-aquatic mammal eDNA signals. eDNA metabarcoding has potential for inclusion in mammal monitoring schemes by enabling large-scale, multi-species distribution assessment for priority and difficult to survey species, and could provide early indication of range expansions or contractions. However, eDNA surveys need high spatiotemporal resolution and metabarcoding biases require further investigation before this tool is routinely implemented.

6.1 Introduction

Globally, mammals are one of the most threatened vertebrate groups, particularly those used for food and medicine, with 23% of species at risk of extinction (Stuart *et al.*, 2004; Butchart *et al.*, 2010). In the UK, many mammal species are under threat due to habitat degradation and loss, non-native species (i.e. competition, hybridisation, disease transmission), or perception as pests (Battersby & Tracking Mammals Partnership, 2005; Massimino *et al*., 2018). The paucity of data for UK terrestrial mammals prevents robust estimation of range expansions or declines and population trends. The majority of species lack long-term, systematic monitoring, and survey effort is particularly biased towards rare species, with widespread species receiving less attention (Massimino *et al*., 2018). Consequently, there is a need for effective and evidence-based strategies for mammal conservation and management (Mathews *et al.*, 2018).

Mammals are generally nocturnal and elusive, requiring observational or acoustic methods for species monitoring (Sadlier *et al.*, 2004; McShea *et al.*, 2016). The most accessible, non-invasive observational methods are field signs (Harris & Yalden, 2004; Sadlier *et al.*, 2004) and camera traps (Ahumada, Hurtado & Lizcano, 2013; Rovero *et al.*, 2014; Burton *et al.*, 2015; Cusack *et al.*, 2015; McShea *et al.*, 2016). Camera trapping especially is cost-efficient, standardised, reproducible, and produces data suited to species occupancy modelling (Ahumada *et al.*, 2013; Rovero *et al.*, 2014; Burton *et al.*, 2015; McShea *et al.*, 2016). However, camera traps can only survey a fraction of large, heterogeneous landscapes, and trap placement can substantially influence species detection probabilities and community insights (Glen *et al.*, 2013; Burton *et al.*, 2015; Cusack *et al.*, 2015; Ishige *et al.*, 2017). Small species in particular are often missed by this approach (Glen *et al.*, 2013; Ishige *et al.*, 2017; Stat *et al.*, 2018). Surveys for field signs are similarly inexpensive, but depend heavily on volunteers with different levels of expertise in order to cover broad geographic areas (Sadlier *et al.*, 2004). Some species also have similar footprints and scat, increasing the potential for misidentification (Harris & Yalden, 2004). The optimal mammal observation method is species-specific, and multiple methods are necessary for large-scale, multi-species monitoring schemes (Battersby & Greenwood, 2004).

Environmental DNA (eDNA) analysis is recognised as a tool for rapid, non-invasive, cost-efficient biodiversity assessment. Organisms transfer their genetic material to their environment via secretions, excretions, gametes, blood, or decomposition, which

can then be isolated from environmental samples (Rees *et al.*, 2014b; Lawson Handley, 2015; Thomsen & Willerslev, 2015). Targeted surveys for single species can be achieved using PCR, quantitative PCR (qPCR), or droplet digital PCR (ddPCR) of eDNA samples. Alternatively, entire communities can be screened using eDNA metabarcoding, where PCR and High-Throughput Sequencing are combined for eDNA analysis (Taberlet *et al.*, 2012; Lawson Handley, 2015; Thomsen & Willerslev, 2015; Deiner *et al.*, 2017). Use of eDNA analysis for single mammal species is increasing, for example, targeted assays are available for cetaceans (Foote *et al.*, 2012; Baker *et al.*, 2018a; Parsons *et al.*, 2018; Qu & Stewart, 2019), manatees (Hunter *et al.*, 2018), platypus (*Ornithorhynchus anatinus*, Lugg *et al.* 2017), otters (Thomsen *et al.*, 2012; Padgett-Stewart *et al.*, 2016), aye-aye (*Daubentonia madagascariensis*, Aylward *et al.*, 2018), and wild boar (*Sus scrofa*, Williams *et al.*, 2018). In contrast, eDNA metabarcoding assessments of mammal communities are rare (Ushio *et al.*, 2017; Klymus *et al.*, 2017b). Mammal assemblages have been obtained from invertebrate blood meals (Schnell *et al.*, 2012; Calvignac-Spencer *et al.*, 2013; Lee, Sing & Wilson, 2015; Tessler *et al.*, 2018) and salt licks (Ishige *et al.*, 2017) in tropical habitats, but samples from the physical environment have tremendous potential to reveal mammal biodiversity over broad spatial and temporal scales (Ushio *et al.*, 2017).

In aquatic ecosystems, eDNA metabarcoding has predominantly been applied to detection of fish (Hänfling *et al.*, 2016; Valentini *et al.*, 2016; Evans *et al.*, 2017a; Lawson Handley *et al.*, 2018), amphibians (Lacoursière-Roussel *et al.*, 2016a; Lopes *et al.*, 2016; Valentini *et al.*, 2016; Sasso *et al.*, 2017), reptiles (Lacoursière-Roussel *et al.*, 2016a), and waterfowl (Ushio *et al.*, 2018b). However, mammals also leave eDNA signatures in water that are distinguishable by metabarcoding (Kelly *et al.*, 2014; Cannon *et al.*, 2016; Hänfling *et al.*, 2016; Port *et al.*, 2016; Andruszkiewicz *et al.*, 2017; Craine *et al.*, 2017; Klymus *et al.*, 2017b; Ushio *et al.*, 2017; Harper *et al.*, 2018b). Ponds in particular act as stepping stones for semi-aquatic and terrestrial taxa (De Meester *et al.*, 2005) by providing opportunities for drinking, foraging, dispersal, and reproduction (Biggs *et al.*, 2016; Klymus *et al.*, 2017b), and could supply natural samples of biodiversity in the wider environment (Deiner *et al.*, 2017; Harper *et al.*, 2018b, 2019a).

Despite evidence to support eDNA deposition in freshwater bodies by semi-aquatic and terrestrial mammals (Thomsen *et al.*, 2012; Rodgers & Mock, 2015; Padgett-Stewart *et al.*, 2016; Klymus *et al.*, 2017b; Lugg *et al.* 2017; Ushio *et al.*, 2017; Harper *et al.*, 2018b; Williams *et al.*, 2018), little is known about the influence of mammal

behaviour on the distribution and strength of the eDNA signal left behind. The most detailed investigation of this nature focused on *S. scrofa*, where limited contact with water was sufficient for eDNA detection, and eDNA from a group of *S. scrofa* remained detectable longer than eDNA from a single individual (Williams *et al.*, 2018). Drinking is a major source of eDNA deposition in water due to saliva, but mammals may also swim, wallow, urinate, or defecate in water (Rodgers & Mock, 2015; Ushio *et al.*, 2017; Williams *et al.*, 2018). Type and frequency of behaviours are extremely diverse across mammal species due to variable lifestyles. For example, arboreal mammals may be less likely to use ponds than semi-aquatic and ground-dwelling counterparts, non-territorial mammals may visit ponds less than those that hold territories, and species that live in groups may leave more DNA in water than those that are solitary (Williams *et al.*, 2018).

In this study, we used two experiments to evaluate eDNA metabarcoding of pond water as a tool for monitoring nine mammal species of conservation or management concern in the UK. The first experiment was designed to examine the role of sampling strategy, mammal lifestyle, and mammal behaviour on eDNA detection and concentration. At two wildlife parks that housed focal species, we employed water sampling at specific sites where behaviours were observed ('directed' sampling) and at equidistant intervals ('stratified' sampling) around artificial water bodies. The second experiment aimed to validate eDNA metabarcoding *in situ*. We sampled water from natural ponds in parallel with camera trapping and field sign searches at sites where focal species were confirmed as present. Our hypotheses for Experiment 1 were as follows: (1) directed sampling would yield stronger eDNA signals for species than stratified sampling; (2) semi-aquatic species would have stronger eDNA signals than ground-dwelling or arboreal species; and (3) behaviours involving mammal contact with water (e.g. swimming, drinking) would be associated with stronger eDNA signals. For Experiment 2, our hypotheses were: (1) eDNA metabarcoding would perform better than camera trapping or field signs for mammal detection; (2) semi-aquatic species eDNA would be evenly distributed in natural ponds, but terrestrial species eDNA would be locally distributed; and (3) mammal eDNA would be detectable for short time frames in comparison to fully aquatic vertebrates.

6.2 Materials and methods

6.2.1 Study species

We studied nine mammal species that are the focus of European conservation, or management: European water vole (*Arvicola amphibius*), European otter (*Lutra lutra*), European beaver (*Castor fiber*), European hedgehog (*Erinaceus europaeus*), European badger (*Meles meles*), red deer (*Cervus elaphus*), Eurasian lynx (*Lynx lynx*), red squirrel (*Sciurus vulgaris*), and European pine marten (*Martes martes*). *A. amphibius*, *L. lutra*, *S. vulgaris*, *M. martes*, and *E. europaeus* are all UK Biodiversity Action Plan (BAP) species (Joint Nature Conservation Committee, 2018). Details on monitoring schemes and population trends for each species are available in reports commissioned by public bodies (Battersby & Partnership, 2005; Mathews *et al.*, 2018), excluding *L. lynx*. *L. lynx* is not currently present in the UK, but reintroduction trials have been proposed (Lynx UK Trust, 2018). *A. amphibius*, *L. lutra* and *C. fiber* are semi-aquatic, *S. vulgaris* and *M. martes* are arboreal, and other species are ground-dwelling. *M. meles* and *C. elaphus* live in groups, whereas other species are solitary except when courting, mating, and rearing young.

6.2.2 Experiment 1: eDNA detection and signal strength in artificial systems

We performed eDNA metabarcoding for focal species under controlled conditions at two wildlife parks in the UK in order to investigate whether the strength of mammal eDNA signals depends on sampling strategy, species lifestyle (semi-aquatic, terrestrial, arboreal), or specific or generic behaviours. Behavioural observation and eDNA sampling were conducted between 18th - 21st September 2017 at Wildwood Trust, Kent, England, and 10th - 11th October 2017 at Royal Zoological Society of Scotland (RZSS) Highland Wildlife Park, Kingussie, Scotland. Sixteen categories of behaviour were defined based on potential contact with water bodies and species lifestyle, and the frequency and duration of these behaviours recorded (Table S6.1). The number of individuals in each enclosure was also recorded alongside size of water bodies (Table 6.1). *C. fiber*, *L. lynx*, *C. elaphus*, and *S. vulgaris* were present at both wildlife parks, whereas other species were only present at Wildwood Trust. Each species was observed for one hour on two separate occasions with exceptions. *M. meles* and *C. fiber* are nocturnal and were observed overnight using camera traps. Behavioural observation was not undertaken for

A. amphibius at Wildwood Trust as animals were under quarantine, or *S. vulgaris* at RZSS Highland Wildlife Park as individuals were wild and widely distributed.

Table 6.1: Summary of focal species studied at each wildlife park and their lifestyle. The number of individuals present and size of waterbodies in a given enclosure is also provided.

Site	Species	Lifestyle	Enclosure	Number of individuals	Water body size (m^2)
Wildwood Trust	European otter (*Lutra lutra*)	Semi-aquatic	1	2	162
	European water vole (*Arvicola amphibius*)	Semi-aquatic	1	4	0.02
			2	1	0.02
	European beaver (*Castor fiber*)	Semi-aquatic	1	2	100
			2	1	100
	European hedgehog (*Erinaceus europaeus*)	Ground-dwelling	1	1	0.04
			2	2	0.04
	European badger (*Meles meles*)	Ground-dwelling	1	4	1.73
	Red deer (*Cervus elaphus*)	Ground-dwelling	1	8	100
	Eurasian lynx (*Lynx lynx*)	Ground-dwelling	1	2	2
	Red squirrel (*Sciurus vulgaris*)	Arboreal	1	2	0.01
			2	3	0.01
			3	3	0.01
			4	2	0.01
	European pine marten (*Martes martes*)	Arboreal	1	1	2
			2	1	0.375
Highland Wildlife Park	Red squirrel (*Sciurus vulgaris*)	Arboreal	NA	NA	0.25
	Eurasian lynx (*Lynx lynx*)	Ground-dwelling	1	8	2
	European beaver (*Castor fiber*)	Semi-aquatic	1	2	50
	Red deer (*Cervus elaphus*)	Ground-dwelling	1	30	NA

Samples were collected from enclosures within 24 hours of behavioural observation. Up to six directed or stratified samples were collected, but the number of samples varied by species due to waterbody size as well as type and frequency of behaviours observed (Tables S6.1, S6.2). If enclosures contained drinking bowls or troughs as well as water bodies, these were also sampled and classed as 'other' rather than directed or stratified samples. No water bodies were present in the *A. amphibius*, *S. vulgaris*, or *E. europaeus* enclosures at Wildwood Trust, thus only drinking bowls were sampled. At RZSS Highland Wildlife Park, the *C. fiber* enclosure had been empty for 24 hours prior to sampling as animals had been moved to RZSS Edinburgh Zoo, Edinburgh, Scotland, for quarantine before wild release. Water from the empty enclosure at RZSS Highland Wildlife Park was sampled nonetheless, and a sample obtained from the *C. fiber* quarantine enclosure at RZSS Edinburgh Zoo. At RZSS Highland Wildlife Park, a sample was also collected from a water bath situated in the woods of the wildlife park to capture any *S. vulgaris* present and classed as 'other'.

In each enclosure, directed samples were collected before stratified samples to minimise disturbance to the water column and risk of cross-contamination. Directed samples were 2 L surface water taken approximately where behaviours were observed. Stratified samples were 2 L surface water (comprised of 8 x 250 mL subsamples) taken at equidistant points around the waterbody perimeter where access permitted. All samples were collected using sterile Gosselin™ HDPE plastic bottles (Fisher Scientific UK Ltd, UK) and disposable gloves. For each species, a field blank (1 L molecular grade water) was taken into the field, opened, then transported alongside samples. Samples collected from Wildwood Trust were transported in sterile coolboxes with ice packs to the University of Kent, where ice was added to coolboxes. Samples were then vacuum-filtered within 6 hours of collection in a wet laboratory that housed exotic amphibians (see Table S6.3), where all surfaces had been sterilised with 10% v/v chlorine-based commercial bleach (Elliott Hygiene Ltd, UK) solution. Samples collected from RZSS Highland Wildlife Park were transported in sterile coolboxes with ice packs to RZSS Edinburgh Zoo, where ice was again added to coolboxes. These samples were vacuum-filtered within 24 hours of collection in a staff meeting room, where all surfaces had been sterilised with 10% v/v chlorine-based commercial bleach (Elliott Hygiene Ltd, UK) solution. All filtration equipment was sterilised before, during, and after set-up in temporary work areas, and upon return to the University of Hull eDNA facility, which is devoted to pre-PCR processes with separate rooms for filtration, DNA extraction and

PCR preparation of environmental samples. Non-electrical equipment was immersed in 10% v/v chlorine-based commercial bleach (Elliott Hygiene Ltd, UK) solution for 10 minutes, followed by 5% MicroSol detergent (Anachem, UK), and rinsed with purified water. Vacuum pumps were wiped with 10% v/v chlorine-based commercial bleach (Elliott Hygiene Ltd, UK) solution.

Where possible, 500 mL of each sample was vacuum-filtered through sterile 0.45 μm cellulose nitrate membrane filters with pads (47 mm diameter; Whatman, GE Healthcare, UK) using Nalgene filtration units. One hour was allowed for each sample to filter, but a second filter was used if filters clogged during this time. A filtration blank (1 L molecular grade water) was also processed during each round of filtration. After 500 mL had been filtered or one hour had passed, filters were removed from pads using sterile tweezers and placed in sterile 47 mm petri dishes (Fisher Scientific UK Ltd, UK), sealed with parafilm (Sigma-Aldrich Company Ltd, UK), and stored at -20 °C in a sterile portable freezer for transport from temporary work areas to the University of Hull. The total volume of water filtered per sample was recorded for downstream analysis (Appendix 6, Table S6.2). After each round of filtration (nine samples/blanks), all filtration units were sterilised in 10% v/v chlorine-based commercial bleach (Elliott Hygiene Ltd, UK) solution for 10 minutes, immersed in 5% v/v MicroSol detergent (Anachem, UK), and rinsed with purified water.

6.2.3 Experiment 2: eDNA detection and signal strength in natural systems

We performed eDNA metabarcoding for focal species under natural conditions at ponds across the UK in conjunction with conventional surveys to validate this molecular approach for mammal identification and to investigate the spatiotemporal variation in mammal eDNA signals. We selected two ponds each at three sites where focal species were confirmed as present. We selected ponds at Bamff Estate, Blairgowrie, Scotland, for *C. fiber*, *L. lutra*, *A. amphibius*, *M. meles*, *C. elaphus*, and *S. vulgaris*, but roe deer (*Capreolus capreolus*) and red fox (*Vulpes vulpes*) were also present. *L. lutra*, *A. amphibius*, and *M. meles* were also present at Tophill Low Nature Reserve, Driffield, East Yorkshire, where American mink (*Neovison vison*), stoat (*Mustela erminea*), weasel (*Mustela nivalis*), rabbit (*Oryctolagus cuniculus*), brown hare (*Lepus europaeus*), *V. vulpes*, *C. capreolus*, and grey squirrel (*Sciurus carolinensis*) additionally occur. We selected Thorne Moors, Doncaster, South Yorkshire, for *C. elaphus* and *M. meles*, but *M.*

erminea, *M. nivalis*, *V. vulpes*, *C. capreolus*, and Reeve's muntjac (*Muntiacus reevesi*) were also present. Camera traps were deployed at Thorne Moors (one at each pond) and Bamff Estate (three at each pond) one week prior to eDNA sampling, and collected once eDNA sampling was complete. At Tophill Low Nature Reserve, camera traps (two at one pond and three at the other pond) were deployed one day before a 5-day period of eDNA sampling every 24 hrs, and collected one week after eDNA sampling was completed. All camera traps were placed so that the pond shoreline and water were in the field of view. Camera traps were set to take three photographs (5 megapixel) when triggered at high sensitivity, with a 3 s interval between triggers.

Stratified samples (10 x 2 L surface water, each comprised of 8 x 250 mL subsamples) were collected from the shoreline of each pond at equidistant points using sterile Gosselin™ HDPE plastic bottles (Fisher Scientific UK Ltd, UK) and disposable gloves. For each pond, a field blank (1 L molecular grade water) was taken into the field, opened, then transported alongside samples. Ponds at Thorne Moors were sampled on 17th April 2018, followed by ponds at Bamff Estate on 20th April 2018. Ponds at Tophill Low Nature Reserve were sampled every 24 hours between 23rd - 27th April 2018. Daily sampling was used to investigate spatiotemporal variation in mammal eDNA signals. Samples collected at Thorne Moors and Tophill Low Nature Reserve were transported on ice in sterile coolboxes with ice packs to the University of Hull eDNA facility, stored at 4 °C, and vacuum-filtered within 6 hours of collection. Samples collected at Bamff Estate were transported in sterile coolboxes with ice packs to local accommodation.

Surfaces and filtration equipment were sterilised before, during, and after set-up in temporary work areas as well as upon return to the University of Hull eDNA facility as in Experiment 1. Samples were vacuum-filtered within 4 hours of collection as outlined in Experiment 1 with minor modifications as follows. Where possible, the full 2 L of each sample were vacuum-filtered, two filters were used for each sample, and duplicate filters were stored in one petri dish. A filtration blank (1 L molecular grade water) was processed during each round of filtration. The total volume of water filtered per sample was recorded (Table S6.4). Filters from Yorkshire sites were immediately stored at -20 °C in a static freezer at the University of Hull, whereas filters from Bamff Estate ponds were transported in a sterile portable freezer (-20 °C) to the static freezer at the University of Hull.

6.2.4 DNA extraction

All DNA was extracted within 2 weeks of filtration at the University of Hull eDNA facility using the water variant of mu-DNA, a protocol tailored for complex environmental samples (Sellers *et al.*, 2018). For Experiment 1, duplicate filters from the same sample were lysed independently and the lysate from each loaded onto one spin column. Due to a larger sample size and number of duplicate filters in Experiment 2, duplicate filters from the same sample were co-extracted by placing both filters in a single tube for bead milling. An extraction blank, consisting only of extraction buffers, was included for each round of DNA extraction (23 samples/blanks). Eluted DNA (100 μL) was stored at -20 °C until PCR amplification.

6.2.5 eDNA metabarcoding

Our eDNA metabarcoding workflow is fully described in Appendix 6. Briefly, we performed a nested metabarcoding workflow that uses a two-step PCR protocol, where Multiplex Identification (MID) tags were included in the first and second PCR for sample identification (Kitson *et al.*, 2019). In the first PCR, eDNA was amplified with published 12S ribosomal RNA (rRNA) primers 12S-V5-F (5'-ACTGGGATTAGATACCCC-3') and 12S-V5-R (5'-TAGAACAGGCTCCTCTAG-3') (Riaz *et al.*, 2011) that were validated *in silico* for all UK vertebrates by Harper *et al*. (2018a, b). These were modified to include MID tags, heterogeneity spacers, sequencing primers, and pre-adapters. PCR positive controls (two per PCR plate; $N = 16$) were exotic cichlid (*Maylandia zebra*) DNA (0.05 ng/μL), and PCR negative controls (two per PCR plate; $N = 16$) were molecular grade sterile water (Fisher Scientific UK Ltd, UK). The first PCR was performed in triplicate for each eDNA sample, and triplicates pooled prior to normalisation. Subsequent PCR products were then pooled according to band strength (see Fig. S6.1) and PCR plate to create sub-libraries for purification with Mag-BIND® RxnPure Plus magnetic beads (Omega Bio-tek Inc, GA, USA), following a double size selection protocol (Bronner *et al.*, 2009).

The second PCR bound pre-adapters, MID tags, and Illumina adapters to the purified sub-libraries. Duplicates for each sub-library were pooled and purified using magnetic beads, following the double size selection protocol (Bronner *et al.*, 2009). Sub-libraries were quantified on a Qubit™ 3.0 fluorometer using a Qubit™ dsDNA HS Assay

Kit (Invitrogen, UK) and pooled proportional to sample size and concentration. The pooled library was purified with magnetic beads using the same ratios, volumes, and protocol as second PCR purification. The library was diluted for quantification by real-time quantitative PCR (qPCR) using the NEBNext® Library Quant Kit for Illumina® (New England Biolabs® Inc., MA, USA). We verified fragment size (330 bp) and removal of secondary product from the library using an Agilent 2200 TapeStation and High Sensitivity D1000 ScreenTape (Agilent Technologies, CA, USA). The library was sequenced at 11.5 pM with 10% PhiX Control on an Illumina MiSeq® using 2 x 300 bp V3 chemistry (Illumina Inc., CA, USA) at the Centre for Ecology & Hydrology, Wallingford.

Raw sequence reads were demultiplexed using a custom Python script, then processed using metaBEAT (metaBarcoding and Environmental Analysis Tool) v0.97.11 (https://github.com/HullUni-bioinformatics/metaBEAT). After quality trimming, merging, chimera detection, and clustering, non-redundant query sequences were compared against our UK vertebrate reference database (Harper *et al.*, 2018a, b) using BLAST (Zhang *et al.*, 2000). Putative taxonomic identity was assigned using a lowest common ancestor (LCA) approach based on the top 10% BLAST matches for any query that matched a reference sequence across more than 80% of its length at a minimum identity of 98%. Unassigned sequences were subjected to a separate BLAST search against the complete NCBI nucleotide (nt) database at 98% identity to determine the source via LCA as described above. The bioinformatic analysis has been deposited in the GitHub repository for reproducibility. The dedicated GitHub repository for this chapter has been permanently archived at: https://doi.org/10.5281/zenodo.2634215.

6.2.6 Data analysis

Analyses were performed in the statistical programming environment R v.3.4.3 (R Core Team, 2017). Data and R scripts have been deposited in the GitHub repository. The total unrefined read counts per sample were calculated and retained for downstream analyses. Assignments from different databases were merged, and spurious assignments (i.e. non-UK species, invertebrates and bacteria) removed from the dataset. The family Cichlidae was reassigned to *Maylandia zebra*. The genera *Bison*, *Bos*, *Buteo*, *Castor*, *Meleagris*, *Pelophylax*, *Sprattus*, *Strix*, and *Triturus* were reassigned to European bison (*Bison bonasus*), cow (*Bos taurus*), common buzzard (*Buteo buteo*), Eurasian beaver, marsh frog

(*Pelophylax ridibundus*), turkey (*Meleagris gallopavo*), European sprat (*Sprattus sprattus*), tawny owl (*Strix aluco*), and great crested newt (*Triturus cristatus*) respectively based on local knowledge of sampling sites and UK distribution maps (National Biodiversity Network Atlas, 2019). The species *Sus scrofa* and *Canis lupus* were reassigned to pig (*Sus scrofa domesticus*) and (*Canis lupus familiaris*) given the restricted distribution of wild boar (*S. scrofa*) and absence of grey wolf (*C. lupus*) in the UK.

Misassignments included the cichlids *Haplochromis burtoni*, *Oreochromis niloticus*, and *Pundamilia nyererei* which were reassigned to *M. zebra*, and Iberian lynx (*Lynx pardinus*) which was reassigned to Eurasian lynx. Other potential misassignments were green-winged teal (*Anas carolinensis*), yellow-browed bunting (*Emberiza chrysophrys*), and Iceland gull (*Larus glaucoides*). These are rare migrants that have been infrequently recorded in the UK (British Trust for Ornithology, 2019) but may have been assigned due to high similarity across reference sequences for different species within the genus *Anas*, and missing reference sequences for several common species within the genera *Emberiza* and *Larus*. These species were reassigned to the genera *Anas*, *Emberiza*, and *Larus*. Reads from corrected assignments were then merged with unaltered assignments.

Of 89 process controls included throughout the metabarcoding workflow, 39 produced no reads. Reads generated for 50 of 89 process controls ranged from 3 to 4930, and strength of each contaminant varied (mean = 62.4%, range = 0.3 - 100.0% of the total reads per process control). Environmental contamination was observed in the field blanks (*M. meles*, *C. fiber*, *L. lynx*, *M. martes*, *S. vulgaris*, and *A. amphibius*) as well as environmental and/or laboratory contamination in the filtration and extraction blanks (human [*Homo sapiens*] and *M. zebra*). PCR negative controls were also contaminated with *H. sapiens*, *M. zebra*, *C. fiber*, and *M. martes* as well as non-focal species (Fig. S3). Consequently, we evaluated different sequence thresholds to minimise the risk of false positives in our dataset. These included the maximum sequence frequency of *M. zebra* DNA in eDNA samples (0.308%), maximum sequence frequency of any DNA except *M. zebra* in PCR positive controls (0.064%), and taxon-specific thresholds (maximum sequence frequency of each taxon in PCR positive controls). The different thresholds were applied to the eDNA samples and the results from each compared (Fig. S4). The taxon-specific thresholds (Table S4) retained the most biological information, thus these were selected for downstream analysis. Consequently, taxa were only classed as present at sites if their sequence frequency exceeded taxon-specific thresholds.

Contaminants remaining in eDNA samples after threshold application included Gentoo penguin (*Pygoscelis papua*) and reindeer (*Rangifer tarandus*) which were likely sourced from the environment, *M. zebra* sourced from the laboratory, and *H. sapiens* which may have originated from the environment or the laboratory. *P. papua* was only detected in water from the beaver quarantine enclosure at RZSS Edinburgh Zoo, and *R. tarandus* was only detected in water sampled from a red squirrel enclosure at Wildwood Trust. *H. sapiens* DNA was detected in the majority of eDNA samples, and *M. zebra* DNA was also present at low frequency in some samples. These contaminants and assignments higher than species level were removed from the dataset, excluding the genera *Anas*, *Emberiza*, and *Larus*. Therefore, all taxonomic assignments in the final dataset were predominantly of species resolution and considered real detections. *B. bonasus*, which is present in the *C. elaphus* enclosure at RZSS Highland Wildlife Park, was detected in two samples taken from this enclosure. However, these detections were excluded from downstream analyses as *B. bonasus* was not one of our focal species. Samples belonging to focal species in Experiment 1 were contaminated with DNA of other focal species to different extents (mean = 6.7%, range = 0.0 - 100.0% of the total refined reads per sample). Therefore, any proportional reads for incorrect focal species in each enclosure were set to 0 for the purposes of downstream analysis.

We subsetted the metabarcoding data according to experiment to generate separate datasets for eDNA samples from artificial water bodies at wildlife parks (Experiment 1) and eDNA samples from natural ponds (Experiment 2). Proportional read counts for each species were calculated from the total unrefined read counts per sample. Samples belonging to focal species in Experiment 1 were contaminated with DNA of other focal species to different extents (mean = 6.7%, range = 0.0 - 100.0% of the total refined reads per sample). Therefore, any proportional reads for incorrect focal species in each enclosure were set to 0 for the purposes of downstream analysis. Our proportional read count data were not normally distributed (Shapiro–Wilk normality test: $W = 0.915$, $P < 0.001$), thus we used a Mann-Whitney *U* test to test for a difference in the median proportional read count of stratified and directed samples across species.

We then employed binomial Generalized Linear Mixed Models (GLMMs) with the logit link function from the development version of the R package glmmTMB (Brooks *et al.*, 2017) to test different hypotheses. First, we examined differences in the eDNA signals produced by stratified and directed samples for each mammal species, with directed samples expected to yield higher proportional read counts than stratified samples.

A hierarchical model, including sample type nested within species as a fixed effect and wildlife park as a random effect, was used. We then tested the hypothesis that species lifestyle would influence mammal eDNA signals, with semi-aquatic species having higher proportional read counts than ground-dwelling or arboreal species. This model included species lifestyle as a fixed effect and species nested within wildlife park as a random effect. Using the directed samples, we tested the hypothesis that mammal behaviour influences their eDNA signals. We expected that behaviours involving direct contact with water (e.g. swimming, drinking) would be associated with higher proportional read counts. We used two hierarchical models that included species nested within wildlife park as a random effect to test the effect of behaviour on proportional read counts. The first modelled specific behaviours as a fixed effect, whereas the second modelled generic behaviour, i.e. water contact versus no water contact. We did not have enough data on behavioural frequencies or duration for different mammal species to test for an effect on proportional read counts. Validation checks were performed to ensure all model assumptions were met where possible and absence of overdispersion (Zuur *et al.*, 2009). Model fit was assessed visually and with the Hosmer and Lemeshow Goodness of Fit Test (Hosmer & Lemeshow, 2000) using the R package ResourceSelection v0.3-2 (Lele *et al.*, 2016). Predictions for each model were obtained using the *predict* function and upper and lower 95% CIs were calculated from the standard error of the predictions. Plots were produced using the R package ggplot2 v3.0.0 (Wickham, 2016).

For Experiment 2, we tested two hypotheses relating to spatiotemporal variation in mammal eDNA signals at natural ponds. First, we qualitatively compared presence-absence records for focal and non-focal mammal species generated by eDNA metabarcoding, camera trapping, and field sign survey, with the expectation eDNA metabarcoding would identify an equivalent or greater number of species than other survey methods. Samples from Tophill Low Nature Reserve spanned a 5-day period, where 10 samples were taken from the same locations in each pond every 24 hrs. The proportional read count data for samples from these ponds were averaged to condense 50 samples to 10 samples per pond for the comparison of method performance across our three study sites. Using the unaveraged data for ponds at Tophill Low Nature Reserve, we then tested the hypothesis that eDNA signals from aquatic species would be more evenly distributed and persist longer than those from terrestrial mammals. We qualitatively assessed change in proportional read counts for identified species over 5 days.

6.3 Results

6.3.1 eDNA metabarcoding

A total of 220 eDNA samples, 89 blanks (field/filtration/extraction), and 32 PCR controls were sequenced. The sequencing run generated 47,713,656 raw sequence reads. A total of 37,590,828 sequences remained following trimming, merging, and application of a length filter. After removal of chimeras and redundancy via clustering, the library contained 21,127,061 sequences (average read count of 91,064 per sample). From these sequences, 16,787,750 (79.46%) were assigned to a taxonomic rank, but 4,339,311 were not assigned a taxonomic identity (20.54%). The final dataset (assignments corrected and thresholds applied) contained 62 vertebrate species (Table S6.6), including six amphibians, 10 fish, 22 birds, and 24 mammals.

6.3.2 Experiment 1: eDNA detection and signal strength in artificial systems

All focal species were detected in all samples taken from water sources in their respective enclosures, excluding *C. elaphus* and *E. europaeus*. *C. elaphus* was not detected in 2 of 20 samples (2 of 5 stratified samples from RZSS Highland Wildlife Park). *E. europaeus* was not detected in 1 of 2 drinking bowl samples (Fig. 6.1). Samples classed as other were excluded from further comparisons, thus *A. amphibius*, *E. europaeus*, and *S. vulgaris* were not represented in downstream analyses. The median proportional read count for stratified samples was 0.406 compared to 0.373 for directed samples across all species, and this difference was not significant (Mann-Whitney U test: U = 1181.5, P = 0.829). Overall, sample type nested within species did not have an effect (χ^2_6 = 0.364, P = 0.999) on proportional read counts (GLMM: θ = 0.168, χ^2_{53} = 8.915, P = 1.000, pseudo-R^2 = 39.21%). Proportional read counts for *L. lutra* were lower than other species, but these differences were not significant (Fig. 6.2a). Similarly, species lifestyle had no influence (χ^2_2 = 0.655, P = 0.721) on proportional read counts (GLMM: θ = 0.213, χ^2_{61} = 13.002, P = 1.000, pseudo-R^2 = 11.85%). Semi-aquatic species had lower and higher proportional read counts than arboreal (-0.491 ± 1.132, Z = -0.434, P = 0.900) and ground-

dwelling species (0.360 ± 0.744, $Z = 0.484$, $P = 0.877$) respectively, whereas proportional read counts for ground-dwelling species were lower than arboreal species (-0.850 ± 1.107, $Z = -0.768$, $P = 0.718$). However, none of these differences were significant (Fig. 6.2b).

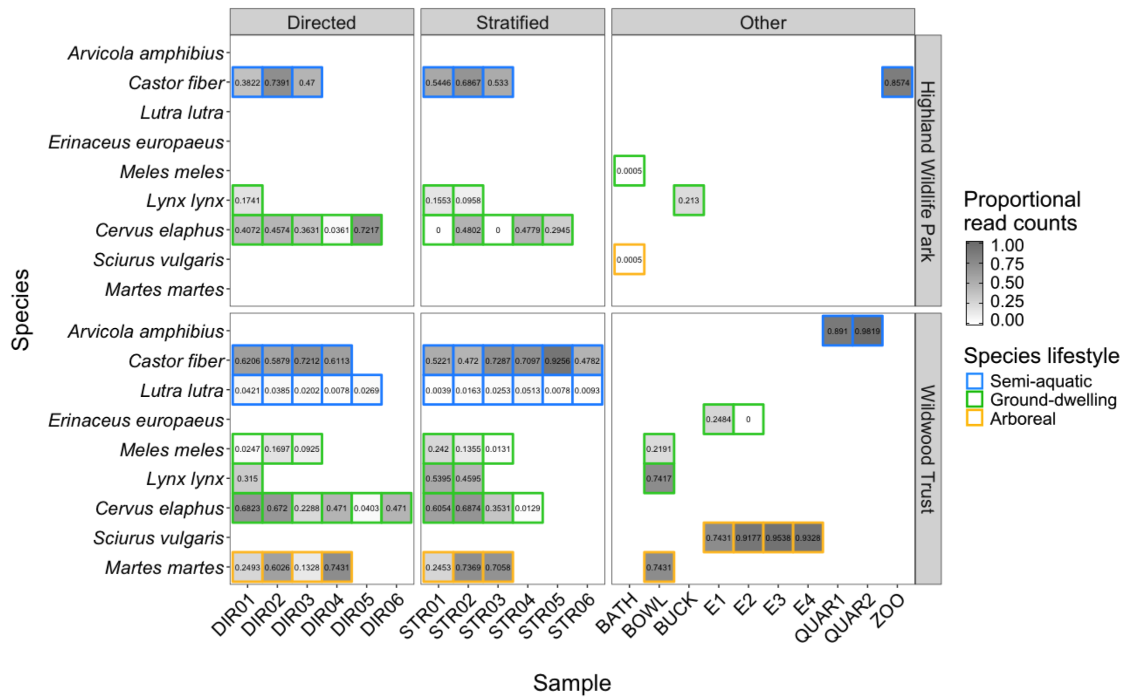


Figure 6.1: Heatmap showing proportional read counts for samples collected from focal species enclosures at wildlife parks. Each square represents a sample that was taken from the enclosure of a particular species. Directed (DIR01-DIR06) or stratified (STR01-STR06) samples were collected from artificial water bodies in species enclosures. Samples were also collected from other sources of water: drinking containers (E1, E2, E3, E4, BOWL, BUCK), quarantine enclosures for water vole (QUAR1, QUAR2) and beaver (ZOO), and a water bath in the woods of RZSS Highland Wildlife Park (BATH).

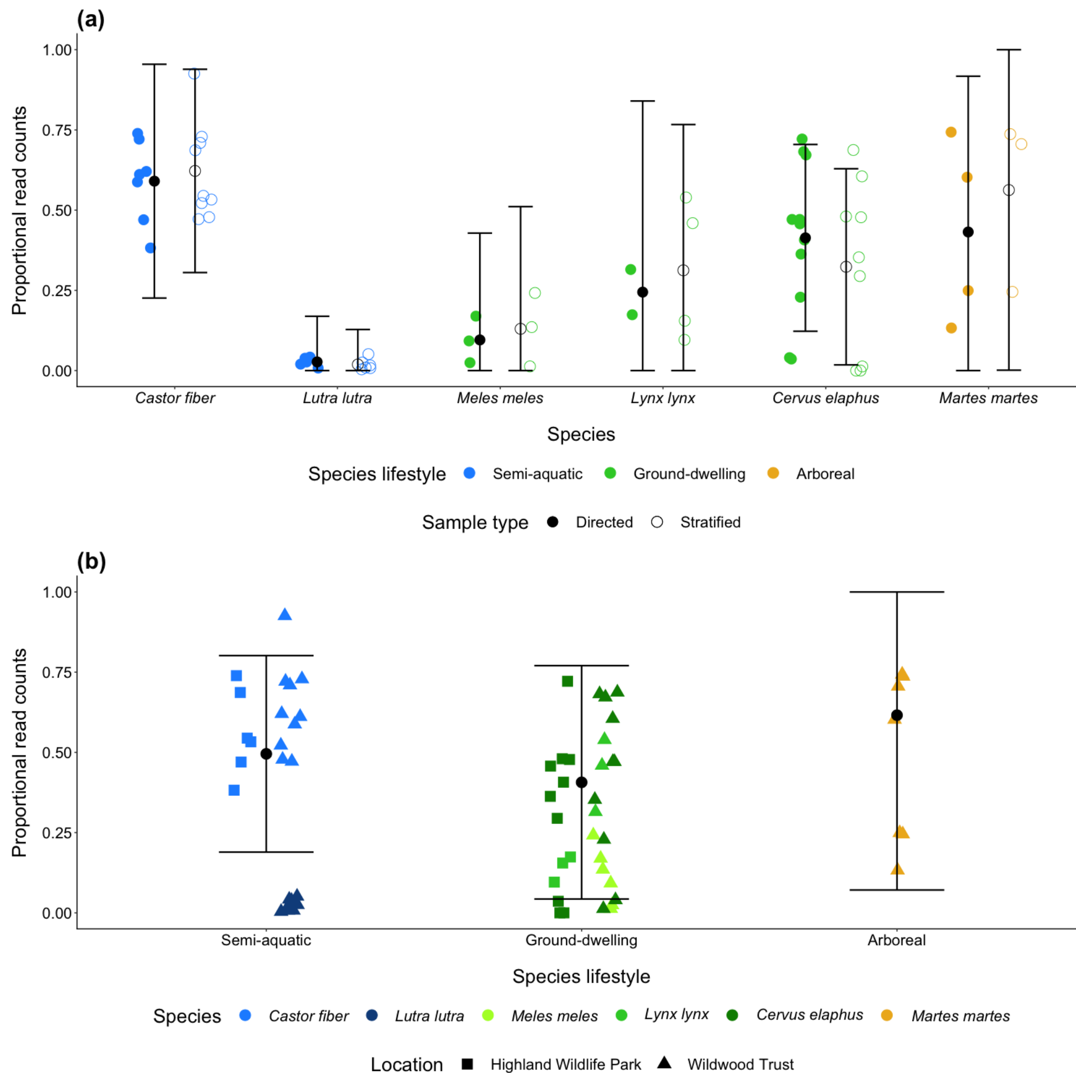


Figure 6.2: Relationships predicted by the binomial GLMMs between proportional read counts and sample type nested within species (a) or species lifestyle (b). The observed data (coloured points) are displayed against the predicted relationships (black points with error bars) for each species **(a)** or species lifestyle **(b)**. Points are shaped by sample type **(a)** or wildlife park **(b)**, and coloured by species. Error bars represent the standard error around the predicted means.

Specific mammal behaviours also had no influence (χ^2_{11} = 1.369, P = 0.999) on proportional read counts (GLMM: θ = 0.355, χ^2_{31} = 11.013, P = 0.999, pseudo-R^2 = 9.17%). Although proportional read counts for most behaviours were lower than proportional read counts for swimming or urination (Fig. 6.3a), these differences were not significant. When specific behaviours were grouped into generic categories, no effect

on proportional read counts was found (χ^2_{11} = 0.002, P = 0.964). Proportional read counts did not differ between behaviour that involved water contact or did not involve water contact (Fig. 6.3b, GLMM: θ = 0.217, χ^2_{41} = 8.897, P = 1.000, pseudo-R^2 = 8.50%).

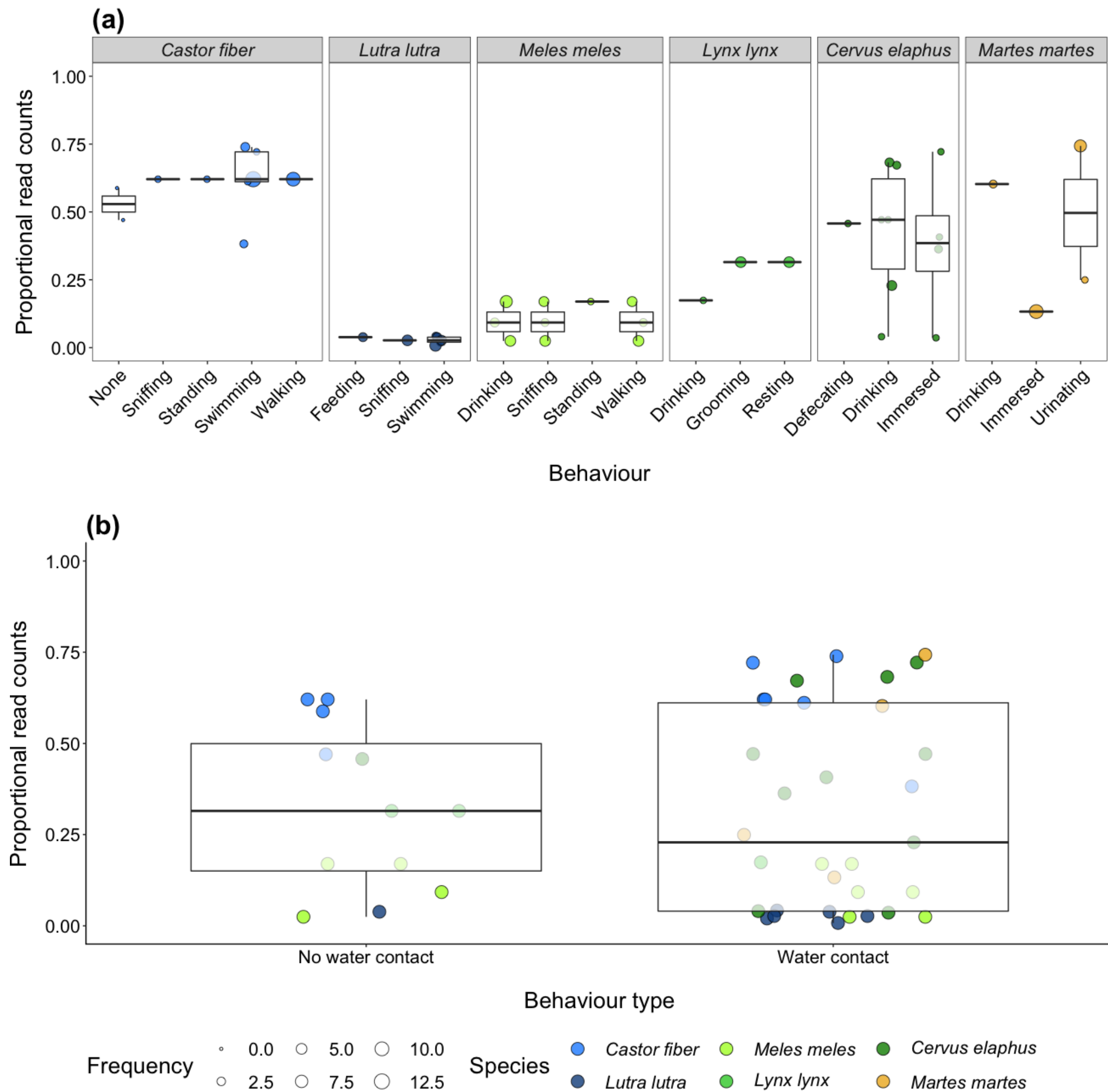


Figure 6.3: Boxplots showing the mean proportional read counts for specific behaviours exhibited by different focal species (a) and behaviour type (b). Boxes show 25th, 50th, and 75th percentiles, and whiskers show 5th and 95th percentiles. Each point represents a directed sample sized by frequency of behaviour in **(a)**. The behaviour 'none' for beaver (*Castor fiber*) represents occurrences of *C. fiber* in water but out of view of camera traps. Beaver and pine marten (*Martes martes*) were the most active species, and also exhibited behaviours in or near water **(a)**. There was no difference in proportional read counts between behaviour involving water contact or no water contact **(b)**.

6.3.3 Experiment 2: eDNA detection and signal strength in natural systems

At natural ponds, eDNA metabarcoding, camera trapping, and field signs all detected *C. elaphus* and *C. fiber* as well as the non-focal *C. capreolus*. Camera traps (Fig. 6.4) and field signs recorded *M. meles* and *V. vulpes* when eDNA metabarcoding did not (Fig. 6.5). However, eDNA metabarcoding revealed several small mammal species not caught on camera, including *A. amphibius* and the non-focal water shrew (*Neomys fodiens*), bank vole (*Myodes glareolus*), common shrew (*Sorex araneus*), brown rat (*Rattus norvegicus*), grey squirrel (*Sciurus carolinensis*), common pipistrelle (*Pipistrellus pipistrellus*), and rabbit (*Oryctolagus cuniculus*). Prints from mice or voles were also observed at Bamff Estate Pond 1, but species could not be determined. Figure 6.5 summarises mammal species recorded by each method at each site with reference to cumulative survey data. Notably, only *C. fiber* presence was captured at the same ponds by all three methods. Although methods shared species at site level, species were not always detected at the same pond. Detection rates for species captured by at least one survey method are summarised in Table S6.7.

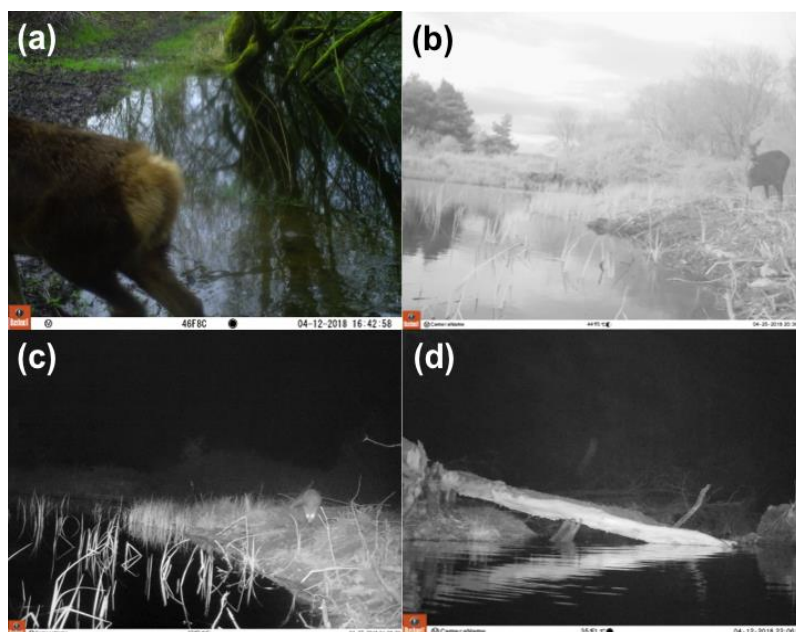


Figure 6.4: Camera trap photographs taken at natural ponds where focal species were confirmed as present. Red deer (*Cervus elaphus*) was recorded at Thorne Moors **(a)**, roe deer (*Capreolus capreolus*) **(b)** and red fox (*Vulpes vulpes*) **(c)** were recorded at Tophill Low Nature Reserve, and beaver (*Castor fiber*) was recorded at Bamff Estate **(d)**.

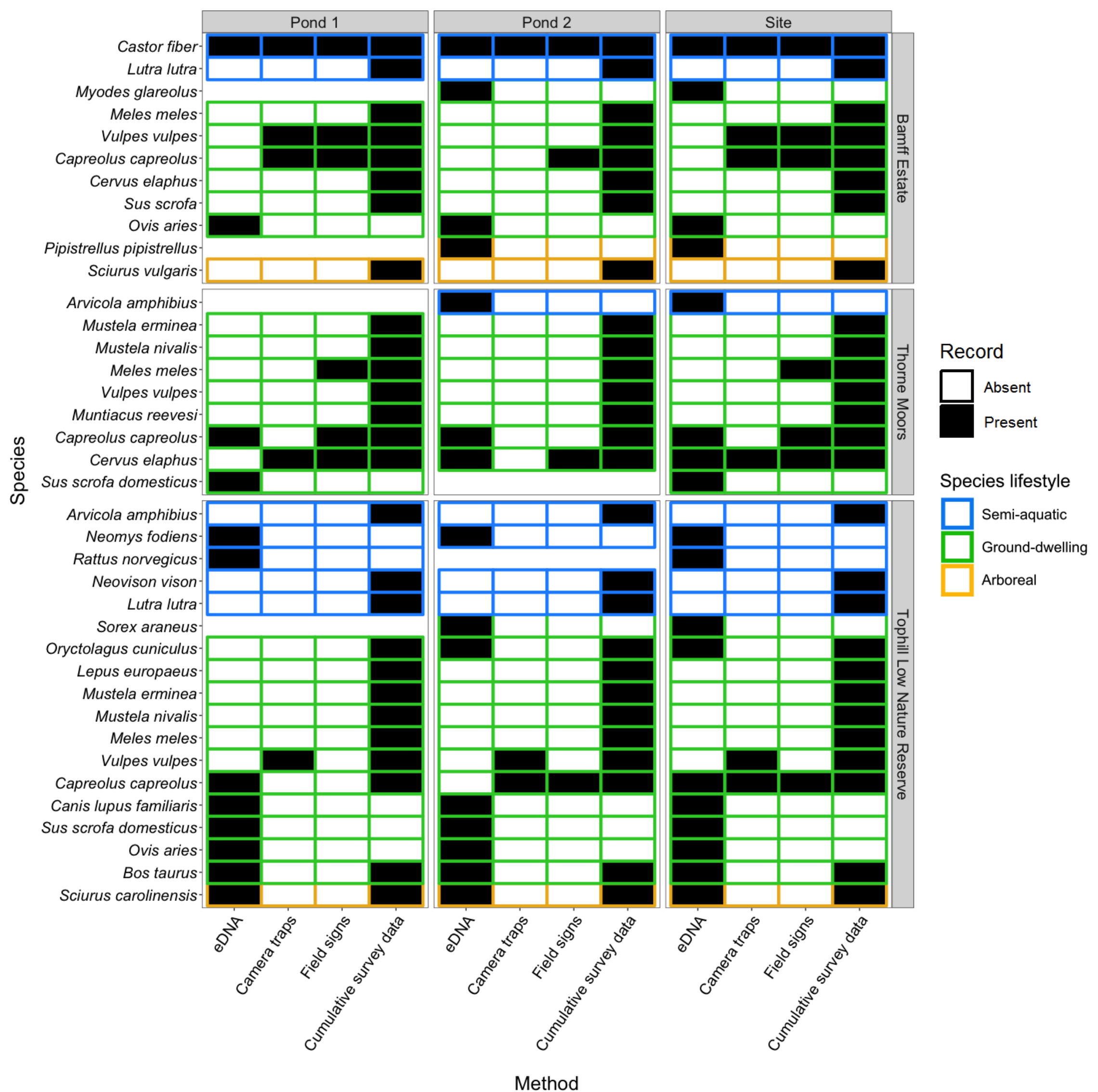


Figure 6.5: Tile plot showing species presence-absence at natural ponds as indicated by field signs, camera trapping, and eDNA metabarcoding at sites where focal species presence was confirmed. Only beaver (*Castor fiber*) was detected by all methods from the same ponds at Bamff estate. Red deer (*Cervus elaphus*) was detected at Thorne Moors by all methods, but not at the same ponds. Similarly, roe deer (*Capreolus capreolus*) was detected at Tophill Low Nature Reserve by all methods, but not at the same ponds.

Sampling of natural ponds revealed spatial and temporal patterns in eDNA detection and signal strength. Considering only mammals, eDNA from terrestrial species was unevenly dispersed compared with semi-aquatic species in natural ponds (Fig. 6.6). The semi-aquatic *A. amphibius* and *C. fiber* were detected in at least 90% and 60% respectively of

water samples ($N = 10$) collected from a given pond, albeit *N. fodiens* was only detected in 10% of samples. Furthermore, eDNA signals from the larger *C. fiber* were highly concentrated. In contrast, non-domestic terrestrial mammals were consistently detected in less than 20% of water samples collected from a pond and left relatively weak eDNA signals. Indeed, eDNA signals from most vertebrate species were unevenly distributed and weak in comparison to those from aquatic amphibians (Fig. 6.6). Analysis of eDNA samples collected over a 5-day period (D01-05) at Tophill Low Nature Reserve revealed that metabarcoding detection of mammals is highly dependent on the spatial and temporal resolution of eDNA surveys. Mammal eDNA signals in pond water were fleeting, often disappearing within 24-48 hrs of initial detection, as opposed to amphibians that were detected for multiple days and whose eDNA signal accumulated in strength. The majority of semi-aquatic or terrestrial species were only detected in a single sample on each day (Fig. 6.7).

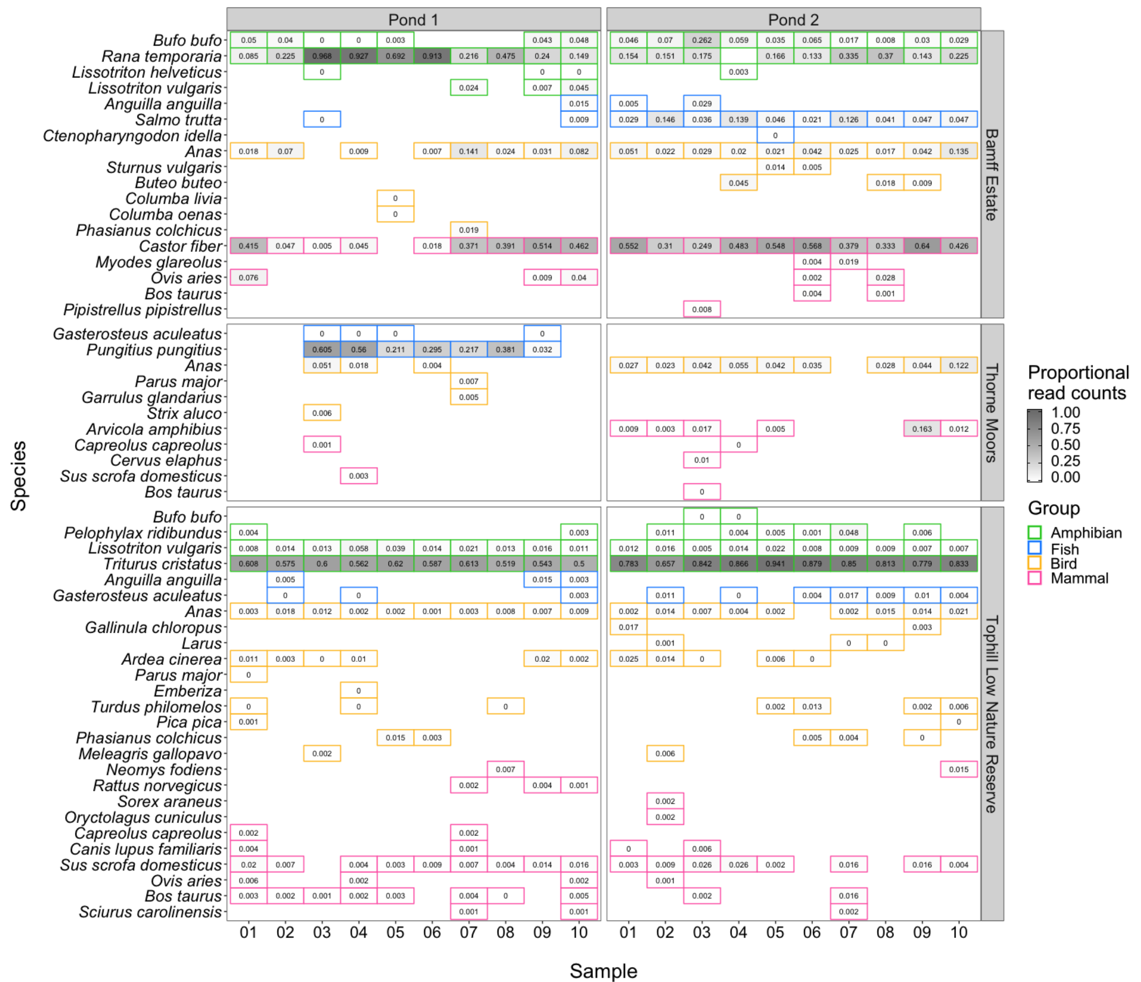


Figure 6.6: Heatmap showing proportional read counts for samples collected from natural ponds at sites where focal species presence was confirmed. Each square represents a sample that had reads assigned to a particular vertebrate species. Species with low proportional read counts (i.e. more than 3 decimal places) are labeled 0. eDNA from semi-aquatic mammals, such as beaver (*Castor fiber*) and water vole (*Arvicola amphibius*), was more concentrated and evenly distributed within ponds than eDNA from terrestrial mammals, e.g. red deer (*Cervus elaphus*) and roe deer (*Capreolus capreolus*). Nonetheless, eDNA from semi-aquatic mammals was less concentrated than eDNA from amphibians, such as great crested newt (*Triturus cristatus*) and common frog (*Rana temporaria*).

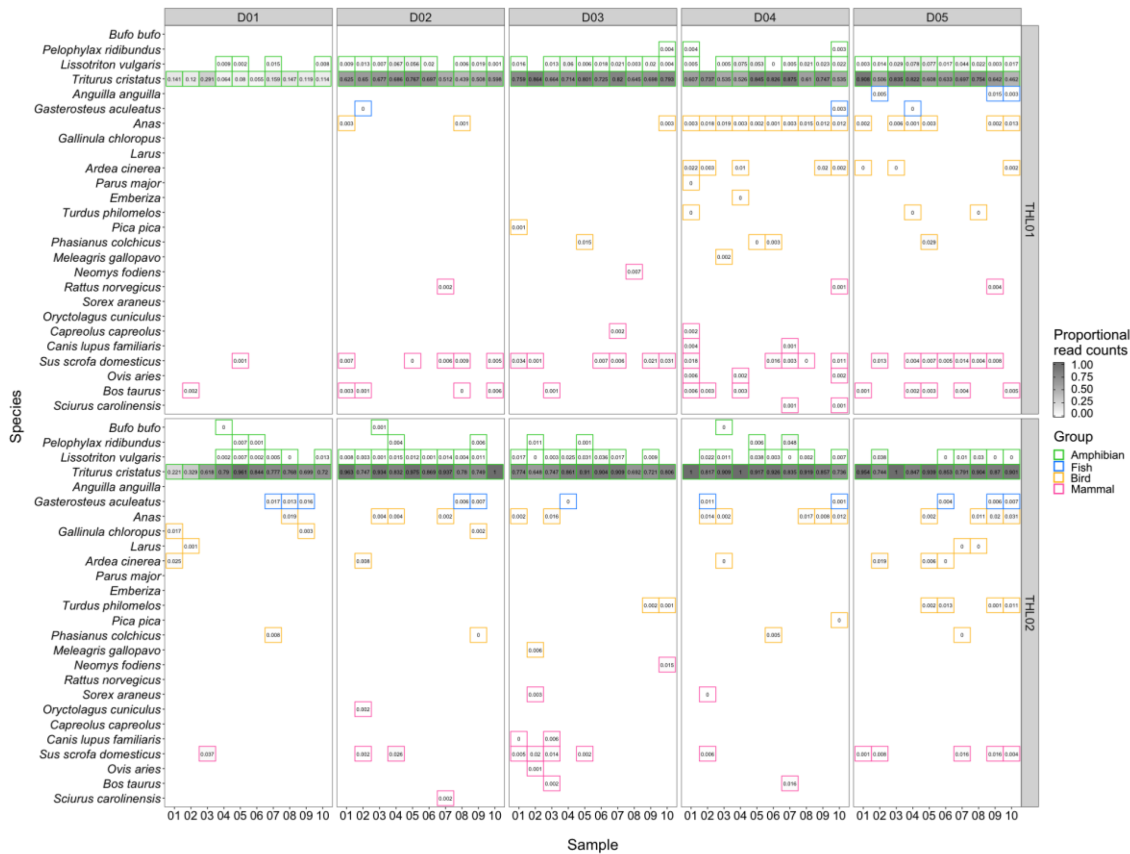


Figure 6.7: Heatmap showing species detected from samples collected at ponds (THL01 and THL02) within Tophill Low Nature Reserve every 24 hrs over a 5-day period (D01 - D05). Each square represents a sample that had reads assigned to a particular vertebrate species. Species with low proportional read counts (i.e. more than 3 decimal places) are labeled 0.

6.4 Discussion

We have explored the use of eDNA metabarcoding as a viable monitoring tool for mammals of conservation and management concern. We used two experiments to validate this molecular approach, and gained new insights that will inform the development and application of eDNA metabarcoding for mammals. Sampling strategy, mammal lifestyle, and mammal behaviour had little influence on eDNA detection and signal strength in captivity, but all played vital roles in natural ponds. Although mammals were detected from pond water, their eDNA signals were ephemeral and weak in comparison to amphibians. Nonetheless, this would suggest eDNA is representative of contemporary and local mammal diversity in a given area.

6.4.1 Experimental insights

In Experiment 1, eDNA detection was achieved from all designated drinking sources for captive mammals, excluding a drinking bowl in one of two *E. europaeus* enclosures. We found no significant differences in eDNA detection or signal strength between stratified and directed samples collected from artificial water bodies across or within species. There were no differences in eDNA signal strength between semi-aquatic, ground-dwelling, and arboreal species. No relationships between proportional read counts and specific or generic behaviours were found. This included those typically associated with eDNA deposition, such as swimming, drinking, urination, and defecation (Rodgers & Mock, 2015; Ushio *et al.*, 2017; Williams *et al.*, 2018). Critically, our experimental set-up must be taken into account when interpreting these results. Artificial water bodies in mammal enclosures were likely saturated with eDNA as individuals had been present for an undetermined length of time prior to sampling. This may have concealed any behavioural relationships. Our sample size was also constrained in that most mammal species were housed at Wildwood Trust only. This resulted in a lack of replication and low experimental power, which may have prevented patterns from being detected statistically. Nonetheless, our results have demonstrated that mammal contact with water will result in eDNA deposition and detection.

Experiment 2 painted a different picture of mammal eDNA detection and signal strength. eDNA metabarcoding successfully detected focal and non-focal mammal species from natural pond water, but detection was not always achieved from each pond at the same site. Field signs and camera trapping performed better than eDNA metabarcoding for *M. meles* and *V. vulpes* detection. Yet, eDNA metabarcoding was the only survey method that identified *A. amphibius* as well as other small mammals not caught on camera or with ambiguous field signs, e.g. mice, voles, shrews. Our findings echo Ishige *et al*. (2017) who achieved mammal detection at salt licks with eDNA metabarcoding comparable to camera trapping, but neither method consistently revealed mammal presence at each salt lick surveyed. Similarly, Klymus *et al*. (2017) did not find the same mammal species utilising a set of uranium mine containment ponds within a site using eDNA metabarcoding. At present, there are no published studies comparing eDNA metabarcoding to camera trapping for mammal identification. However, a study on marine fish biodiversity found species richness was highest using both baited remote

underwater video systems and eDNA metabarcoding, but unique species were identified by each method (Stat *et al.*, 2018).

Notably, no survey method captured *L. lutra* presence. *L. lutra* was successfully detected in eDNA metabarcoding studies of UK ponds (Harper *et al.*, 2018b) and lakes (Hänfling *et al.*, 2017) as well as in Experiment 1 here. Nonetheless, we observed a weaker eDNA signal for *L. lutra* in comparison to other semi-aquatic mammals in captivity. eDNA metabarcoding may have performed poorly for *L. lutra*, *M. meles*, and *V. vulpes* due to the ecology of these species. These mammals are wide-ranging (Thomsen *et al.*, 2012; Gaughran *et al.*, 2018) and may not readily release DNA in water. The otter tends to spraint on grass or rock substrata outside water and use latrines associated with caves and dens (Ruiz-Olmo & Gosálbez, 1997). Similar to other terrestrial mammals, eDNA deposition by *M. meles* and *V. vulpes* will depend on these species drinking from or entering ponds (Rodgers & Mock, 2015; Ushio *et al.*, 2017; Williams *et al.*, 2018). Therefore, eDNA detection of *L. lutra*, *M. meles*, and *V. vulpes* may require greater temporal and spatial resolution of water sampling. Alternatively, false negatives may be indicative of underlying issues with the metabarcoding approach (see section 6.4.2).

Within natural ponds, eDNA from semi-aquatic mammals (e.g. *A. amphibius*, *C. fiber*) tended to have an even distribution, being found in all or most samples collected on fine spatial scales, whereas eDNA from terrestrial mammals (e.g. *C. capreolus*, *C. elaphus*) was highly localised and often detected in less than 20% of samples. Mammal eDNA signals also varied temporally, with detection achieved for a maximum of two consecutive days. Mammal eDNA may be spatially and temporally clumped in lentic ecosystems due to the nature and frequency of contact that individuals have with water. Mammals may only be detected at drinking sites unless they exhibit behaviours that involve prolonged water contact, such as swimming and wallowing (Klymus *et al.*, 2017b; Ushio *et al.*, 2017; Williams *et al.*, 2018). eDNA detection and persistence are further influenced by group size, where eDNA from multiple individuals endures for longer periods in water than eDNA from single individuals (Williams *et al.*, 2018). Therefore, detailed investigations into the density of individuals in a given area that incorporate biotic (e.g. sex, body mass, behaviour) and abiotic (e.g. temperature, pH, ultraviolet light) factors are needed to understand the longevity of mammal eDNA signals in aquatic ecosystems (Rodgers & Mock, 2015; Parsons *et al.*, 2018; Williams *et al.*, 2018).

Both of our experiments have shown that sampling strategy can drastically

influence eDNA detection probability for mammals. In captivity, mammal eDNA was evenly distributed in closed, artificial water bodies, whereas in open, natural ponds mammal eDNA was found to be locally distributed. At each wildlife park, mammal species were housed individually (excluding *C. elaphus* at RZSS Highland Wildlife Park) with a designated drinking source and/or small water body (range 0.01 - 162 m^2, mean 27.4 m^2) in each enclosure. Additionally, some enclosures housed more individuals of a species than others, thereby increasing eDNA deposition rate and probability of detection (Williams *et al.*, 2018). In contrast, wild mammals have an array of freshwater habitats at their disposal and can hold vast territories, thus rates of pond visitation and eDNA deposition are more random and irregular (Klymus *et al.*, 2017b; Ushio *et al.*, 2017) which may lead to between-sample variation (Williams *et al.*, 2018).

Where small ponds are being studied in areas with dense mammal populations, probability of eDNA detection will likely be high. However, rigorous sampling strategies are required to track mammals in areas with large water bodies that are sparsely populated by mammals. Importantly, we sampled natural ponds in spring, but sampling in other seasons may produce different results. For example, an eDNA metabarcoding study of fish biodiversity in Windermere found high repeatability in eDNA detection and abundance estimates across seasons, but strong spatial differences reflective of species ecology (Lawson Handley *et al.*, 2018). Therefore, we recommend that researchers and practitioners using eDNA metabarcoding for mammal monitoring channel their efforts into extensive sampling of numerous water bodies in a given area over larger timescales. This will account for differential mammal visitation rates and maximise probability of eDNA detection.

6.4.2 Pitfalls of eDNA metabarcoding for mammal monitoring

eDNA metabarcoding has strong potential for inclusion in mammal monitoring schemes (see section 6.4.3) but akin to existing survey tools, it suffers from biases that may result in false negatives or false positives. The most important of these when targeting mammals is contamination (Cannon *et al.*, 2016; Port *et al.*, 2016; Klymus *et al.*, 2017b; Ushio *et al.*, 2017; Egeter *et al.*, 2018). Our process controls identified low-level contamination from domestic and wild species at all stages of metabarcoding, but contaminants primarily occurred in the field blanks or PCR negative controls. We also identified *M. zebra* sequences in the field, filtration, and extraction blanks, even though this DNA was

not handled prior to PCR. Therefore, contaminants found in process blanks and PCR negative controls may result from PCR contamination (Kelly *et al.*, 2014) or sequencing error (Hänfling *et al.*, 2016). Indeed, PCR negative controls corresponding to samples from Tophill Low Nature Reserve were contaminated with great crested newt (*Triturus cristatus*), a highly abundant species in ponds at the nature reserve. This would imply that highly concentrated DNA and eDNA samples can contaminate negative controls during metabarcoding. Although negative controls at each stage of the metabarcoding workflow can help identify when contaminants were introduced (Cannon *et al.*, 2016; Ushio *et al.*, 2017; Klymus *et al.*, 2017b), contaminants in these controls can amplify exponentially with no competition affecting inferences (Harper *et al.*, 2018a). Therefore, innovative approaches are needed to minimise and mitigate contamination in metabarcoding. For example, estimation of false positives using site occupancy modelling (Ficetola *et al.*, 2015) or sequencing thresholds, i.e. the number of sequence reads required for a sample to be species positive (Hänfling *et al.*, 2016; Civade *et al.*, 2016; Evans *et al.*, 2017a).

We applied taxon-specific sequence thresholds to our metabarcoding data to minimise false positives. After threshold application, the only contaminants remaining in biological samples were *M. zebra* (laboratory contaminant), *P. papua*, *R. tarandus* (environmental contaminants), and *H. sapiens* (environmental/laboratory contaminant). *P. papua* is housed at RZSS Edinburgh Zoo and was identified from water in the *C. fiber* quarantine enclosure. The *S. vulgaris* and *R. tarandus* enclosures are in close proximity at Wildwood Trust, but not directly next to each other. A possible explanation for this environmental contamination is DNA transport by keepers as they completed their duties. *H. sapiens* DNA was the most severe contaminant, but was likely sourced from the environment rather than the laboratory. Both wildlife parks were open to visitors and had staff working in or around enclosures at time of sampling. At natural sites, *H. sapiens* was also omnipresent, for example, as dog walkers, bird watchers, wildlife photographers, and reserve staff. Unfortunately, sources of environmental contamination cannot be eliminated and have consequences for eDNA metabarcoding (Kelly *et al.*, 2014; Port *et al.*, 2016; Ushio *et al.*, 2017). *H. sapiens* DNA may be amplified and sequenced instead of focal species, potentially resulting in false negative detections for rare and/or less abundant species (Boessenkool *et al.*, 2012; Valentini *et al.*, 2016; Klymus *et al.*, 2017b; Ushio *et al.*, 2017; Egeter *et al.*, 2018). This can be prevented with the use of blocking primers for *H. sapiens* DNA (Boessenkool *et al.*, 2012; Valentini *et al.*, 2016), but blocking primers can also impair amplification of species of interest (Port *et al.*, 2016;

Ushio *et al.*, 2017).

In our study, eDNA metabarcoding produced false negatives for *L. lutra*, *M. meles*, and *V. vulpes* at natural ponds. DNA from aquatic and more abundant species may have overwhelmed *L. lutra*, *M. meles*, and *V. vulpes* DNA during amplification and sequencing, i.e. species-masking (Kelly *et al.*, 2014; Klymus *et al.*, 2017b). More biological and technical replication could help mitigate this amplification bias and improve species detection probabilities (Valentini *et al.*, 2016; Andruszkiewicz *et al.*, 2017; Evans *et al.*, 2017a; Lawson Handley *et al.*, 2018; Stat *et al.*, 2018), but there may also be an issue of primer bias. We selected a universal 12S rRNA primer pair designed to amplify DNA from all vertebrates (Riaz *et al.*, 2011) for metabarcoding. There were no mismatches between the forward or reverse primer and any *L. lutra*, *M. meles*, and *V. vulpes* sequences. During *in silico* tests, all species amplified when up to three mismatches between each primer and reference sequences were allowed. However, some species may have been preferentially amplified *in vitro* due to greater primer binding affinity (Kelly *et al.*, 2014; Andruszkiewicz *et al.*, 2017; Evans *et al.*, 2017a; Klymus *et al.*, 2017b; Stat *et al.*, 2018). Primers designed to target mammals, such as the MiMammal primers from Ushio *et al.* (2017), or multi-marker (e.g. 12S, 16S, COI) investigations (Kelly *et al.*, 2014; Hänfling *et al.*, 2016; Evans *et al.*, 2017a; Klymus *et al.*, 2017b) may improve metabarcoding detection of these species in systems where there is more competition from aquatic species. Notably, Thomsen *et al.* (2012) also observed lower qPCR detection for *L. lutra* than fish or amphibians. A comparison of metabarcoding and targeted qPCR for *L. lutra* would test whether low amplification efficiency is due to metabarcoding issues or the ecology of this species. Similar comparisons have been made for the threatened *T. cristatus* (Harper *et al.*, 2018a) and wood turtle (*Glyptemys insculpta*, Lacoursière-Roussel *et al.*, 2016a) with different outcomes. This would confirm whether eDNA metabarcoding can be reliably used to monitor *L. lutra* alongside the wider mammal community.

6.4.3 Scope of eDNA metabarcoding for mammal monitoring

Despite issues inherent to the metabarcoding approach for biodiversity monitoring (Deiner *et al.*, 2017), this tool has enormous potential to enhance mammal monitoring, conservation, and management. The most recent assessment of UK mammal populations emphasised the paucity of data and lack of systematic monitoring for many species

(Mathews *et al.*, 2018). Distribution and occupancy data is poor for most species, with ongoing survey effort biased toward rare species. These surveys are also heavily reliant on citizen science and casual records. Consequently, there is a need for tools that can provide standardised, systematic monitoring of UK mammal populations (Mathews *et al.*, 2018). eDNA metabarcoding generates distribution data for multiple species, whether rare, invasive, or abundant. This tool could be particularly powerful for tracking species in conflict with one another. For example, *A. amphibius*, *L. lutra*, and *N. vison* (Bonesi & Macdonald, 2004), or *M. martes*, *S. carolinensis*, and *S. vulgaris* (Sheehy *et al.*, 2018).

eDNA metabarcoding can rapidly (sampling to sequencing in three weeks) survey multitudes of sites at large-scales where camera traps would be resource-intensive, cost-inefficient, and susceptible to theft or damage (Glen *et al.*, 2013; Ushio *et al.*, 2017; Stat *et al.*, 2018). Field signs can be employed at comparable spatial scales to eDNA metabarcoding but depend on volunteer time and skill (McShea *et al.*, 2016). Metabarcoding could also provide species resolution data for species that are misidentified from field signs, e.g. mice and voles, *L. lutra* and *N. vison* (Harris & Yalden, 2004). However, these conventional approaches should not be thrown out in favour of eDNA metabarcoding. Both camera traps and field signs recorded species that eDNA metabarcoding did not. Therefore, eDNA metabarcoding is complementary and should be incorporated into existing monitoring schemes. For mammal monitoring, eDNA metabarcoding could be most effective if deployed at the edges of known species distributions, in areas where species presence is unknown, and in areas with isolated species records (Mathews *et al.*, 2018). Different sample types (e.g. water from lotic and lentic ecosystems, soil, snow, salt licks, feeding traces, faeces, and blood meals) may also offer new insights to mammal biodiversity in a given area (Ishige *et al.*, 2017; Ushio *et al.*, 2017; Aylward *et al.*, 2018; Tessler *et al.*, 2018).

6.5 Acknowledgements

This work was funded by the University of Hull. We would like to thank Helen Zenn (RZSS Edinburgh Zoo) for her input on the study design, and Richard Griffiths (DICE, University of Kent) for his support with filtration of water samples from Wildwood Trust. We are grateful to staff at Wildwood Trust and RZSS Highland Wildlife Park for their cooperation and assistance with eDNA sampling from animal enclosures. Richard

Hampshire and volunteers at Tophill Low Nature Reserve also deserve recognition for their assistance with deploying camera traps. Tim Kohler (Natural England) gave permission to sample at Thorne Moors.

6.6 Data accessibility

Raw sequence reads have been archived on the NCBI Sequence Read Archive (Study: SRP164740; BioProject: PRJNA495011; BioSamples: SAMN10195928 - SAMN10196255; SRA accessions: SRR7986451 - SRR7986778). Jupyter notebooks, R scripts and corresponding data are deposited in a dedicated GitHub repository (https://github.com/lrharper1/LRHarper_PhDThesis_Chapter6) which has been permanently archived (https://doi.org/10.5281/zenodo.2634215).

Chapter 7: General Discussion

Grass snake (*Natrix natrix*) (Linnaeus, 1758) swimming through pond
© user: GrahamC57 | Flickr | CC BY-NC-ND 2.0

Certain content in sections 7.1 and 7.4 – 7.5 of this chapter was written for a review paper on eDNA monitoring in ponds that was first-authored by LRH and published in *Hydrobiologia* as

Harper, L.R., Buxton, A.S., Rees, H.C., Bruce, K., Brys, R., Halfmaerten, D., Read, D.S., Watson, H.V., Sayer, C.D., Jones, E.P., Priestley, V., Mächler, E., Múrria, C., Garcés-Pastor, S., Medupin, C., Burgess, K., Benson, G., Boonham, N., Griffiths, R.A., Lawson Handley, L. & Hänfling, B. (2019) Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. *Hydrobiologia*, **826**, 25–41.

7.1. Can eDNA analysis be used to monitor threatened biodiversity associated with ponds?

eDNA analysis has been applied worldwide to survey threatened or rare pond biota, but the extent of its application varies widely by geographic location (Lawson Handley, 2015; Harper *et al.*, 2019a). Rare or threatened taxa associated with ponds that have been studied using targeted or passive eDNA approaches include invertebrates (Thomsen *et al.*, 2012; Doi *et al.*, 2017), amphibians (Thomsen *et al.*, 2012; Valentini *et al.*, 2016; Bálint *et al.*, 2018; Goldberg *et al*., 2018), reptiles (Raemy & Ursenbacher, 2018), fish (Li *et al.*, 2018b, c), birds (Ushio *et al.*, 2018b), and mammals (Thomsen *et al.*, 2012; Ushio *et al.*, 2017). In the UK, targeted eDNA analysis has only been applied to the detection of the great crested newt (*Triturus cristatus*, Rees *et al.*, 2014a; Biggs *et al.*, 2015), invasive (Dunn *et al.*, 2017; Blackman *et al.*, 2018; Harper *et al.*, 2018c; Robinson *et al.*, 2018) or native invertebrates (Robinson *et al.*, 2018; Seymour *et al.*, 2018), and threatened (Seymour *et al.*, 2018) or non-native fish (Davison *et al.*, 2016). In the UK, passive eDNA analysis (i.e. eDNA metabarcoding) has also been used to survey freshwater invertebrate (Bista *et al.*, 2015, 2017; Blackman *et al.*, 2017) and fish (Hänfling *et al.*, 2016; Lawson Handley *et al.*, 2018; Li *et al.*, 2018b, c; Li *et al.*, 2019) assemblages as well as marine eukaryotes (Deiner *et al.*, 2018). The number of available assays and published studies is expected to increase rapidly over the coming years (*pers. comm.* UK DNA Working Group Meeting 2018). However, *T. cristatus* (Rees *et al.*, 2014a, 2017; Biggs *et al.*, 2015; Buxton *et al*., 2017a, b, 2018) and fish (Davison *et al.*, 2016; Li *et al.*, 2018b, c; Harper *et al.*, 2019b) remain the only taxa to have been surveyed in ponds. All other aforementioned taxa were studied in lakes, rivers, or seaports.

This thesis has empirically demonstrated the power and value of eDNA approaches for monitoring threatened amphibians (Chapters 2 & 3), fish (Chapters 3 & 4), and mammals (Chapters 3 & 6) associated with ponds in the UK. In Chapter 2, passive and targeted eDNA approaches with detection thresholds offered comparable detection of the threatened *T. cristatus*. Furthermore, adaptations to the eDNA metabarcoding workflow would likely improve detection sensitivity. In Chapter 4, targeted eDNA survey was comparable to conventional fyke netting for the imperiled crucian carp (*Carassius carassius*). Finally, in Chapter 6, eDNA metabarcoding was complementary to camera trapping and field sign survey for conservation and management priority mammals utilising ponds. Therefore, eDNA analysis has untapped potential to monitor threatened

pond biodiversity that should be exploited and further investigated. In addition to the diagnostic strength and versatile applications of this monitoring tool, general advantages include minimal disturbance to target taxa, weather independence, reduced risk of disease transmission, and time and cost-efficiency (Biggs *et al.*, 2014; Rees *et al.*, 2014a; Tréguier *et al.*, 2014; Valentini *et al.*, 2016; Bálint *et al.*, 2018). Nonetheless, I identified constraints associated with using this tool in pond ecosystems, particularly those discussed in Chapters 2 and 4.

In Chapter 2, both passive and targeted eDNA analysis failed to detect *T. cristatus* in some ponds where eggs were recorded, but egg searches had a greater false negative rate for *T. cristatus* overall. In Chapter 4, I also observed imperfect detection (90%) for *C. carassius* using targeted eDNA analysis. eDNA analysis was only compared to conventional fyke netting, but other fish survey tools are applicable to *C. carassius*, e.g. electrofishing and acoustic telemetry (Hardie *et al*., 2006). In Chapter 6, eDNA metabarcoding failed to detect mammal species that left field signs or were captured on camera, but instead revealed the presence of smaller mammals that leave ambiguous field signs or elude camera traps. False negatives produced by eDNA analysis may stem from the ponds surveyed, or the technical limitations of this method. eDNA is patchily distributed in ponds due to organisms being unevenly distributed (Takahara *et al.*, 2012; Eichmiller *et al*., 2014) across microhabitats that are used for different purposes, i.e. feeding and reproduction (Goldberg *et al.*, 2018). Horizontal eDNA dispersion is restricted by barriers to water movement (Biggs *et al.*, 2015), and vertical eDNA dispersion limited by chemical stratification of the water column due to minimal wind-mixing (Sayer *et al.*, 2013). Consequently, sample number and water volume can have substantial impacts on eDNA detection probability (Schultz & Lance, 2015; Goldberg *et al.*, 2018). For maximal species detection, eDNA sampling in ponds must be exhaustive and cover many different locations over fine spatial scales (Goldberg *et al.*, 2018), encompassing areas around/underneath barriers to water movement and different depths (Harper *et al.*, 2019a). Ecology of the target species, including life stage, condition, seasonality, and behaviour, should always be considered when designing eDNA surveys due to its capacity to influence eDNA detection (Smart *et al.*, 2015; Spear *et al.*, 2015; de Souza *et al.*, 2016; Buxton *et al.*, 2017b, 2018; Rees *et al.*, 2017; Harper *et al.*, 2018c; Takahashi *et al.*, 2018). Downstream, capture and extraction methods dictate the quality and concentration of eDNA. Different workflows are optimal for different target species, thus independent comparisons of capture and extraction methods are needed

(Lacoursière-Roussel *et al*., 2016b; Piggott, 2016; Spens *et al.*, 2016; Hinlo *et al.*, 2017b). Technical replication will also improve detection sensitivity (Schultz & Lance, 2015; Piggott, 2016), but assays must be robust to false positives (Goldberg *et al.*, 2016). Similarly, the possibility of false negatives induced by sample inhibition must be excluded by including synthetic Internal Positive Controls (Goldberg *et al.*, 2016) or exogenous control DNA (Doi *et al.*, 2017) in amplification reactions. Ponds are particularly prone to inhibitor build-up, thus eDNA detection in these ecosystems may suffer without preventive measures (Harper *et al.*, 2019a).

In addition to imperfect detection for *C. carassius*, there was uncertainty around the relationship observed between *C. carassius* density with eDNA detection and quantification. Correlations between conventional and eDNA-based estimates of relative abundance or biomass for target taxa have been found in ponds (Takahara *et al.*, 2012; Thomsen *et al.*, 2012; Biggs *et al.*, 2015; Buxton *et al.*, 2017b), but these are not consistently observed (Rees *et al.*, 2014a; Doi *et al.*, 2017; Raemy & Ursenbacher, 2018). Estimates of relative abundance or biomass produced by conventional survey tools can be unreliable due to biased capture of particular sexes, ages, and size classes by these tools as well as bias induced by season of and time of deployment (Hardie *et al.*, 2006; Ruane *et al.*, 2012; Turner *et al.*, 2012). Before eDNA analysis is deemed incapable of inferring relative abundance or biomass, eDNA survey should occur in close spatial and temporal proximity to multiple conventional survey tools and the estimates produced by all methods compared. However, eDNA-based estimates of relative abundance or biomass are variable in freshwater ecosystems due to the effects exerted by biotic and abiotic factors on eDNA release, persistence, and degradation (Barnes *et al.*, 2014; Strickler *et al*., Goldberg, 2015; Buxton *et al.*, 2017b; Goldberg *et al.*, 2018). Effects of abiotic factors may be more pronounced in ponds that experience environmental extremes not observed in other freshwater habitats, e.g. hydroperiod, nutrient loading, pH, conductivity (De Meester *et al.*, 2005; Goldberg *et al.*, 2018). As a result, relative abundance or biomass estimates may only be feasible when target taxa are at a particular life stage or exhibit certain behaviours, e.g. fertilisation, egg production, spawning (Buxton *et al.*, 2017b; Bylemans *et al.*, 2017; Dunn *et al.*, 2017). In ponds, temperature (Takahara *et al.*, 2012; Robson *et al.*, 2016; Buxton *et al.*, 2017b; Goldberg *et al.*, 2018), pH (Goldberg *et al.*, 2018), conductivity (Harper *et al.*, 2019b), and sediment type (Buxton *et al.*, 2017a) had impacts on target eDNA concentration. Therefore, caution is needed when performing relative abundance or biomass estimation for pond species to

exclude the influences of under-representative sampling, inhibition, and biotic or abiotic variables. eDNA monitoring for threatened pond biota will prosper with further investigations into the role of organisms (e.g. habitat use, species ecology, abundance, biomass) and abiotic variables (e.g. temperature, pH, ultraviolet light, anoxia, conductivity) on eDNA release, persistence, degradation, and detection (Harper *et al.*, 2019a).

A potential issue when monitoring threatened species as part of community surveys using eDNA metabarcoding is species masking, i.e. eDNA signals for organisms that are rare or low density in their environment are overwhelmed by those of more abundant species (Kelly *et al.*, 2014). In Chapter 2, I compared eDNA metabarcoding to qPCR for *T. cristatus* detection to determine whether species masking is problematic for monitoring of threatened pond biodiversity. Despite less PCR replication, detection sensitivity of eDNA metabarcoding with no detection threshold was comparable to qPCR with a detection threshold. eDNA metabarcoding of pond water also revealed 59 other vertebrate species and did not require additional investigator effort or cost to qPCR. These findings highlight that both targeted and passive eDNA approaches must be used with care, and the results of each interpreted with caution. Based on the lower detection rate for *T. cristatus*, eDNA metabarcoding may fail to detect rare species within communities. Biological replication (Andruszkiewicz *et al.*, 2017; Bálint *et al.*, 2018), marker choice (Kelly *et al.*, 2014; Valentini *et al.*, 2016), metabarcode primer design and amplicon length (Lacoursière-Roussel *et al.*, 2016a), technical replication, and sequencing depth (Kelly *et al.*, 2014; Civade *et al.*, 2016; Port *et al.*, 2016) are critical to prevent species masking. Furthermore, eDNA sampling, capture, and extraction strategies will further influence detection of species within a community (Shaw *et al.*, 2016b; Djurhuus *et al.*, 2017; Klymus *et al.*, 2017b; Deiner *et al.*, 2018). qPCR with no detection threshold may generate false positives for rare species, thus detection thresholds for qPCR should be evaluated (Goldberg *et al.*, 2016; Smart *et al.*, 2016). Further comparisons of qPCR and eDNA metabarcoding for rare species detection are required. These should permute sampling (pseudoreplicates from merged sample vs. biological replicates) and eDNA capture (ethanol precipitation vs. filtration) as well as use the same level of PCR/sequencing and qPCR replication to enable a fair comparison of detection sensitivity.

Importantly, in Chapters 2, 3, or 6 of this thesis, no reptiles were detected using eDNA metabarcoding. This included the legally protected grass snake (*Natrix natrix*)

which is highly characteristic of ponds and uses these habitats to regulate temperature and to hunt (Reading & Jofré, 2009). Targeted eDNA analysis has been successful for some species of freshwater turtle and terrestrial snake (Piaggio *et al.*, 2014; Hunter *et al.*, 2015; Lacoursière-Roussel *et al.*, 2016a; Kucherenko *et al.*, 2018), but prone to failure for other species at sites with known presence (Baker *et al.*, 2018b; Raemy & Ursenbacher, 2018). Concerning passive eDNA analysis, Lacoursière-Roussel *et al.* (2016a) found that eDNA metabarcoding had poor performance for reptiles in comparison to amphibians. Similarly, Kelly *et al*. (2014) were unable to detect green sea turtle (*Chelonia mydas*) in a mesocosm experiment with marine fishes. Reptiles may be more challenging to monitor using eDNA analysis due to lower eDNA shedding rates. Freshwater turtles lack epithelial cells or mucus and produce minimal secretions as well as highly concentrated excretions in comparison to amphibians and fish (Raemy & Ursenbacher, 2018). Semi-aquatic snakes spend less time in water, defecate less frequently, exhibit low activity, and tend to shed skin on land (Hunter *et al.*, 2015). Therefore, further investigations are needed to assess the capability of eDNA analysis to monitor reptiles associated with ponds.

Despite the issues identified, I conclude that eDNA analysis could revolutionise monitoring of threatened pond biodiversity. This tool could upscale rare or threatened species monitoring through rapid and cost-efficient screening of multitudes of sites over large spatial and temporal scales (Chapter 1; Harper *et al.*, 2019a). This mass data generation will radically improve distribution mapping and occupancy modelling for rare or threatened species (Thomsen *et al.*, 2012; Biggs *et al.*, 2015; Doi *et al.*, 2017; Goldberg *et al.*, 2018).

7.2 Can eDNA metabarcoding be used to survey biodiversity at the pondscape, including semi-aquatic and terrestrial taxa?

Deiner *et al*. (2016) previously highlighted the potential of rivers to provide catchment-scale information on aquatic and terrestrial biodiversity. Ponds possess the same data mining potential as both permanent aquatic (Valentini *et al.*, 2016; Evans *et al.*, 2017a; Bálint *et al.*, 2018) and visiting terrestrial species (Klymus *et al.*, 2017b; Ushio *et al.*, 2017, 2018b) can be identified from eDNA present in these ecosystems. Therefore, ponds can provide natural samples of biodiversity in the wider environment (Deiner *et al.*, 2017;

Harper *et al.*, 2019a). However, no study to date has launched a holistic assessment of invertebrate and/or vertebrate biodiversity associated with ponds.

This thesis has empirically demonstrated the potential of eDNA metabarcoding to simultaneously survey aquatic, semi-aquatic, and terrestrial taxa linked to ponds, thereby traversing the aquatic-terrestrial boundary in pondscape monitoring and research. More than 150 invertebrate (Chapter 5) and over 60 vertebrate (Chapters 2, 3 & 6) species, including fish, amphibians, birds, and mammals, were detected using eDNA metabarcoding on water collected from different sets of ponds across the UK. Furthermore, passive eDNA survey was complementary to targeted eDNA survey and egg searches for *T. cristatus* (Chapter 2), morphotaxonomic identification and DNA metabarcoding for invertebrates (Chapter 5), and camera trapping and field sign survey for mammals (Chapter 6). Despite the potential of eDNA metabarcoding to generate distribution data *en masse* for species across the tree of life, there are challenges to be overcome in the field, laboratory, and at the keyboard before this tool can be routinely implemented for biodiversity monitoring in any ecosystem (Thomsen & Willerslev, 2015; Deiner *et al.*, 2017).

Detection of species by eDNA metabarcoding and species richness estimates produced by this approach are highly dependent on sampling timeframe and completeness to counter the effects of spatiotemporal dynamics and eDNA transport, whether human or animal-mediated (Deiner *et al.*, 2016; Hänfling *et al.*, 2016; Olds *et al.*, 2016; Andruszkiewicz *et al.*, 2017; Macher & Leese, 2017; Lawson Handley *et al.*, 2018). eDNA metabarcoding may also be unable to provide abundance or biomass estimates for all species within a community. In aquatic ecosystems, relative abundance estimates have been made for fish communities identified using eDNA metabarcoding that are consistent with estimates produced by conventional monitoring tools, e.g. electrofishing, seine netting, gill-netting, trawling (Evans *et al.*, 2017a; Hänfling *et al.*, 2016; Thomsen *et al.*, 2016; Lawson Handley *et al.*, 2018; Ushio *et al.*, 2018a; Li *et al.*, 2019). However, relative abundance estimates for diverse taxonomic assemblages could be confounded by potential amplification bias that occurs during metabarcoding (Chapters 2 & 6 Discussion; Deiner *et al.*, 2017; Klymus *et al.*, 2017b).

Species detection as well as richness, abundance, and biomass estimates by eDNA metabarcoding are likely to be influenced by the ecology of eDNA (Barnes & Turner, 2015), biotic and abiotic factors (Barnes *et al.*, 2014; Strickler *et al.*, 2015; Buxton *et al.*, 2017b; Macher & Leese, 2017; Goldberg *et al.*, 2018), sampling strategy (Macher &

Leese, 2017; Lawson Handley *et al.*, 2018) and biological replication (Andruszkiewicz *et al.*, 2017; Evans *et al.*, 2017a; Bálint *et al.*, 2018), eDNA capture (Djurhuus *et al.*, 2017; Klymus *et al.*, 2017b; Deiner *et al.*, 2018; Li *et al.*, 2018b), eDNA extraction (Djurhuus *et al.*, 2017; Deiner *et al.*, 2018), primer design (Bylemans *et al.*, 2018), technical replication (Kelly *et al.*, 2014; Ficetola *et al.*, 2015; Civade *et al.*, 2016; Port *et al.*, 2016), library preparation (Schnell *et al.*, 2015; O'Donnell *et al.*, 2016; Leray & Knowlton, 2017), and bioinformatic processing (Clare *et al.*, 2016; Evans *et al.*, 2017a). Continued investigations using eDNA metabarcoding in conjunction with multiple conventional tools are required to disentangle the influence of these effects on eDNA in relation to species detection as well as the contemporary and local abundance of species within a community (Deiner *et al.*, 2017). Nevertheless, across ecosystems, studied taxa, and ecological contexts, eDNA metabarcoding often provides biological information that is orders of magnitude greater and more reliable than the information produced by conventional monitoring tools (Hänfling *et al.*, 2016; Shaw *et al.*, 2016a; Valentini *et al.*, 2016; Sasso *et al.*, 2017; Bálint *et al.*, 2018; Lawson Handley *et al.*, 2018; Nakagawa *et al.*, 2018; Stat *et al.*, 2018; Li *et al.*, 2019).

The challenges to achieving reliable estimates for biological communities via eDNA metabarcoding identified above are even more pronounced for invertebrates. Invertebrate diversity has been investigated in rivers (Deiner *et al.*, 2016; Blackman *et al.*, 2017; Klymus *et al.*, 2017a; Macher & Leese, 2017), streams (Macher *et al.*, 2018), and lakes (Bista *et al.*, 2017) using eDNA metabarcoding. Ponds have been neglected in this regard despite the more diverse and unique invertebrate communities they possess in comparison to other freshwater habitats (Williams *et al.*, 2003; Davies *et al.*, 2008). Nonetheless, common themes have emerged from eDNA metabarcoding studies of invertebrates in other freshwater habitats. Metabarcoding for invertebrates currently relies on the cytochrome c oxidase subunit I (*COI*) gene due to its roots in DNA barcoding (Hebert *et al.*, 2003) and the available reference databases for this marker (Curry *et al.*, 2018). However, *COI* metabarcoding primers (Meusnier *et al.*, 2008; Zeale *et al.*, 2011; Geller *et al.*, 2013; Leray *et al.*, 2013; Elbrecht & Leese, 2017) are often degenerate to allow binding at highly variable sites found throughout this protein-coding gene (Deagle *et al.*, 2014; Clarke *et al.*, 2017; Elbrecht *et al.*, 2018). As a result of high degeneracy, these primers can have low and unpredictable amplification efficiency (Deagle *et al.*, 2014), bind non-target regions (Elbrecht *et al.*, 2018), fail to recover all taxa present (Elbrecht *et al.*, 2016), or amplify non-metazoan taxa, e.g. bacteria, fungi, and algae

(Brandon-Mong *et al.*, 2015; Macher & Leese, 2017; Macher *et al.*, 2018). This amplification bias can occur even if primers are designed to target a metazoan group (Elbrecht & Leese, 2017) and has consequences for reliability of taxonomic identifications as well as abundance estimation (Deagle *et al.*, 2014). The use of multiple markers (e.g. *COI*, 16S, 18S) or development of more specific primers can alleviate the problems associated with *COI* (Elbrecht *et al.*, 2016; Klymus *et al.*, 2017a), but reference databases for these alternative markers must be supplemented by researchers and end users invested in DNA-based monitoring tools (Elbrecht *et al.*, 2016; Curry *et al.*, 2018).

Although not explored in this thesis, eDNA analysis has enormous potential for macrophyte survey in ponds (Fujiwara *et al.*, 2016; Matsuhashi *et al.*, 2016; Newton *et al.*, 2016; Gantz *et al.*, 2018; Kuzmina *et al.*, 2018). Like conventional monitoring tools for many other taxa, macrophyte surveys typically entail observation or capture of species and their identification based on morphological features. This is laborious, time-consuming, reliant on taxonomic expertise (Fujiwara *et al.*, 2016; Gantz *et al.*, 2018), and limited by species phenology and microscopic traits (Kuzmina *et al.*, 2018). Targeted eDNA analysis has been found to alleviate the challenges associated with detecting invasive, submerged macrophytes in mesocosms (Scriver *et al.*, 2015; Gantz *et al.*, 2018) as well as natural rivers, lakes (Gantz *et al.*, 2018; Newton *et al.*, 2016), and ponds (Fujiwara *et al.*, 2016; Matsuhashi *et al.*, 2016). However, passive eDNA analysis (i.e. eDNA metabarcoding) could have deeper implications for macrophyte community survey. Kuzmina *et al.* (2018) recently demonstrated the utility of eDNA metabarcoding to monitor pondweeds along a river, where previously documented and new species were detected. Detection of macrophyte communities via eDNA metabarcoding could improve diversity estimates and indicate water quality with applications in biomonitoring. However, marker choice together with complete and accurate reference databases are crucial to successful implementation of this tool (Kuzmina *et al.*, 2018).

My results indicate that eDNA metabarcoding can and should be used to survey aquatic and non-aquatic biodiversity at the pondscape. Ponds have exceptionally high biodiversity value, but monitoring is problematic due to the number of ponds found across landscapes and the limitations of available sampling tools, e.g. taxonomic expertise, under-representation of certain taxa (Biggs *et al.*, 2016; Hill *et al.*, 2018). eDNA metabarcoding could resolve these issues and provide species-level distribution data for entire communities to inform pondscape conservation and management (Harper *et al.*, 2019a).

7.3 What are the prospects of eDNA metabarcoding for community investigation in ponds?

eDNA metabarcoding is recognised as a tool that could induce a step change in freshwater conservation, management, monitoring, and research (Hering *et al.*, 2018). Applications of this tool include multi-species occupancy modelling, species richness and diversity estimation, examination of spatiotemporal variation in biological communities across ecosystems, species and trophic networks, biomonitoring, and citizen science (Deiner *et al.*, 2017; Harper *et al.*, 2019a). However, previous eDNA metabarcoding studies of ponds have tended to focus on the technical aspects of this tool and the species inventories it produced, rather than the ecological applications of the data generated (Valentini *et al.*, 2016; Evans *et al.*, 2017a; Ushio *et al.*, 2017, 2018b; Bálint *et al.*, 2018). This is an area of research that must be addressed to ensure uptake of this tool by end users in routine biodiversity monitoring of ponds. In this thesis, some of these applications were empirically tested. eDNA metabarcoding was used to distinguish biotic and abiotic determinants of *T. cristatus* and vertebrate species richness at the UK pondscape (Chapter 3). In conjunction with morphotaxonomic identification and DNA metabarcoding, eDNA metabarcoding was implemented to assess the impact of fish stocking on alpha and beta diversity of pond invertebrates (Chapter 5). Finally, the spatiotemporal variation in eDNA signals left by semi-aquatic and terrestrial mammals in ponds was examined using eDNA metabarcoding (Chapter 6).

eDNA metabarcoding has tremendous scope and unprecedented diagnostic power to enable hypothesis testing relating to the distribution of biodiversity and its response to environmental pressures at larger spatial and temporal scales (Deiner *et al.*, 2017). This potential has begun to be explored in studies that have estimated species richness and examined diversity along environmental gradients using eDNA metabarcoding (Hänfling *et al.*, 2016; Kelly *et al.*, 2016; Olds *et al.*, 2016; Andruszkiewicz *et al.*, 2017; Evans *et al.*, 2017a; Macher & Leese, 2017; Lawson Handley *et al.*, 2018; Macher *et al.*, 2018). eDNA metabarcoding is particularly applicable to test the effects of environmental gradients in ponds which experience extreme hydroperiod, temperature, and pH (De Meester *et al.*, 2005); however, there are many more insights to be gained. In Chapter 3, I demonstrated the capacity of eDNA analysis to upscale freshwater monitoring and

research, and used *T. cristatus* as a case study to ground truth eDNA metabarcoding for ecological hypothesis testing. I verified biotic and abiotic determinants of *T. cristatus* revealed by eDNA metabarcoding in over 500 ponds against those widely reported in the existing literature on this species. I also explored the potential of eDNA metabarcoding to reveal determinants of vertebrate species richness at the UK pondscape. Many biotic (species associations) and abiotic (physical properties of ponds) determinants revealed by the eDNA metabarcoding data were supported by the existing *T. cristatus* literature. Shared determinants between *T. cristatus* and vertebrate species richness also implied that *T. cristatus* conservation measures would benefit wider biodiversity at the pondscape. This chapter signifies that eDNA metabarcoding is a highly applicable tool to test a range of ecological hypotheses for different taxa.

In Chapter 3, I provided examples of analyses that could be performed to identify determinants of particular species within a community or the community itself; however, there are other avenues available. Biotic interactions can be investigated through ecological network analysis; for example, data generated by DNA metabarcoding was analysed using ecological networks to examine species interactions within a terrestrial ecosystem (Evans *et al.*, 2016a). Occupancy modelling of eDNA data for single species (see Chapter 4) is often undertaken as a vital component of targeted assay validation (Schmidt *et al.*, 2013; Hunter *et al.*, 2015; de Souza *et al.*, 2016; Buxton *et al.*, 2017a; Goldberg *et al.*, 2018; Harper *et al.*, 2019b; Strickland & Roberts, 2019), but adopted to a lesser extent for multi-species eDNA data. Multi-species occupancy modelling could be used to estimate detection probabilities for a variety of species within a community. This analysis was implemented by Valentini *et al*. (2016) to identify factors influencing detection probability for 10 amphibian species identified by eDNA metabarcoding from pond water. Multi-species occupancy modelling is also a useful approach to estimate the number of false positives produced by eDNA metabarcoding (Ficetola *et al.*, 2015; Ficetola, Taberlet & Coissac, 2016).

Richness and diversity estimation are highly applicable to the community data generated by eDNA metabarcoding (Olds *et al.*, 2016; Macher & Leese, 2017; Li *et al.*, 2018a; Macher *et al.*, 2018; Nakagawa *et al.*, 2018). Alpha (site) and beta (between-site) diversity in particular are typically the focus of community ecology investigations. Several studies have estimated alpha and beta diversity from eDNA metabarcoding data, but these focused on lotic ecosystems (Li *et al.*, 2018a; Macher *et al.*, 2018; Nakagawa *et al.*, 2018). At time of writing, there are no studies that have estimated both alpha and beta

diversity of lentic ecosystems based on eDNA metabarcoding. In Chapter 5, I tested another ecological hypothesis by assessing the impact of fish stocking on invertebrate diversity in ponds. I compared alpha and beta diversity of invertebrates in ponds containing *C. carassius* to fishless ponds. Analyses were performed for data generated by morphotaxonomic identification, DNA metabarcoding, eDNA metabarcoding, and all methods combined. The alpha and beta diversity produced by these three methods of invertebrate assessment was then compared. *C. carassius* was found to have a negligible or minor impact on alpha diversity, and a positive (albeit weak or moderate) effect on beta diversity of pond invertebrates. Ponds with *C. carassius* possessed different species and families to fishless ponds, thus *C. carassius* presence resulted in dissimilar community composition across the pond network. Method of invertebrate assessment produced different estimates of alpha diversity, and had a strong, positive effect on community structure. eDNA metabarcoding generated the most species and families for ponds, but also produced a vastly different community to other methods. My results have demonstrated the potential of eDNA metabarcoding for alpha and beta diversity estimation in pond ecosystems, and indicate that multiple methods of invertebrate assessment should be used to best inform freshwater conservation and management.

In Chapter 3 of this thesis, an array of aquatic and non-aquatic biodiversity (60 vertebrate species) was recorded at the UK pondscape. However, without support from historical or conventional data, these eDNA detections could arguably have resulted from eDNA transport between ponds by humans, domestic animals, or waterfowl (Deiner *et al.*, 2017; Harper *et al.*, 2019a). In Chapter 6, I launched a thorough investigation into mammal eDNA signals found in pond water. In captivity, samples directed by mammal behaviour and stratified samples at equidistant intervals around artificial water bodies were taken. The eDNA signal (i.e. proportional read counts) from these were compared, and the directed samples examined in relation to species lifestyle (semi-aquatic, ground-dwelling, and arboreal) and behaviour. At natural ponds, stratified water samples were taken in conjunction with camera trapping and field sign survey. Furthermore, at one site, eDNA sampling was conducted over a 5-day period to investigate spatiotemporal variation in mammal eDNA signals. All mammals were detected in captivity and no significant effects of species lifestyle or behaviour found. eDNA signals from mammals were evenly distributed in artificial water bodies. At natural ponds, eDNA metabarcoding detected three focal species and revealed the presence of species that camera trapping and field signs did not capture; although, eDNA metabarcoding failed to record species that

conventional methods did. eDNA signals from semi-aquatic species were evenly distributed in ponds, but eDNA signals for terrestrial species were highly localised. Moreover, eDNA signals from mammals only persisted for 24-48 hrs during the 5-day sampling period. My results echo those of Ushio *et al*. (2017), and reinforce that eDNA metabarcoding studies of mammals must be conducted over fine spatial and temporal scales to capture all species present in a given area.

eDNA metabarcoding holds promise for ecological research in ponds through upscaled biodiversity monitoring at greater spatial and temporal resolution. This can lead to improved distribution mapping and occupancy modelling, testing of ecological hypotheses, identification of species interactions, richness and diversity estimation, biomonitoring, and investigation of spatiotemporal dynamics. These applications should be fully exploited to enhance understanding of biodiversity associated with individual ponds and pond networks, the aquatic-terrestrial boundary in pondscapes, environmental and anthropogenic stressors of pond communities, and the ecology of eDNA, i.e. the origin, state, fate, and transport of eDNA (Barnes & Turner, 2015), for multiple species simultaneously.

7.4 Overcoming the limitations of eDNA metabarcoding

eDNA metabarcoding has many prospects for community investigation in ponds, but there is much to be done to ensure its accuracy and reliability for routine biodiversity monitoring. Limitations of this technology include incomplete reference databases, metabarcode choice, and reproducibility of bioinformatics and data analysis. Global initiatives such as GenBank (www.ncbi.nlm.nih.gov/genbank) and BOLD (www.boldsystems.org) are in place to improve taxonomic and geographic coverage of reference sequence databases. However, it will take several years to resolve eDNA monitoring issues at lower taxonomic levels (Ficetola *et al*., 2010; Comtet *et al*., 2015; Cowart *et al*., 2015). Metabarcoding markers often differ from the standardised markers that were used to construct reference databases based on morphologically identified specimens. Consequently, there is disparity between conventional and molecular species identification when they should be integrated (Cristescu, 2014). Existing databases are predisposed toward the *COI* region which is substandard for eDNA metabarcoding (Thomsen & Willerslev, 2015) and has associated biases (see Chapter 5 and Deagle *et al*.,

2014). Universal metabarcoding primers based on the mitochondrial 12S, 16S or 18S ribosomal RNA regions could be highly effective for broad biodiversity assessments, but can be limited by database representation (Zinger, Gobet & Pommier, 2012; Deagle *et al*., 2014; Elbrecht *et al*., 2016). Reference databases must supply sequences for various markers to enable accurate species identification and discovery of new species (Elbrecht *et al.*, 2016; Curry *et al.*, 2018). This will require considerable time investment and taxonomic expertise as well as internationally standardised collection and laboratory methods (Cristescu, 2014). However, these efforts are crucial to prevent species misidentifications and false negatives in eDNA metabarcoding assessments (see Chapters 3 and 5).

In addition to procurement and curation of reference sequence databases, the future of eDNA metabarcoding depends on reliability and reproducibility of bioinformatics pipelines. These pipelines use a suite of softwares to merge, filter, remove, cluster and assign taxonomic identities to raw sequence reads. However, the softwares used vary widely across studies, with pipelines making use of functions from several different wrappers or toolkits, such as OBITools, QIIME, USEARCH, VSEARCH, RDP classifier and MOTHUR. The majority of these toolkits rely on UNIX operating systems and must be implemented using the command line, which challenges inexperienced programmers and hinders reproducibility (Coissac *et al*., 2012; Dufresne *et al*., 2019). This thesis used the metaBEAT (https://github.com/HullUni-bioinformatics/metaBEAT) pipeline in conjunction with Jupyter notebooks and R for reproducibility of bioinformatics, data manipulation, and statistical analyses. The metaBEAT pipeline has extensive documentation and implements established softwares for processing eDNA metabarcoding data: trimming using Trimmomatic (Bolger, Lohse & Usadel 2014), merging using FLASH (Magoč & Salzberg 2011), chimera removal using UCHIME algorithm (Edgar *et al.* 2011) from VSEARCH (Rognes *et al.* 2016), clustering using VSEARCH, and taxonomic assignment with BLAST (Zhang *et al.* 2000), although other methods are available (e.g. Kraken, pplacer). The metaBEAT pipeline is not actively maintained thus its longevity cannot be guaranteed. Nonetheless, documented and user-friendly bioinformatics pipelines (e.g. SLIM, insect) are on the rise and hold strong potential to improve reproducibility in eDNA metabarcoding (Wilkinson *et al*., 2018; Dufresne *et al*., 2019).

The aforementioned issues are being tackled by independent research groups, but international initiatives are also underway. In Europe, DNAqua-net is a network of

researchers and end users invested in the development of gold-standard molecular tools and indices for biodiversity assessment and biomonitoring of water bodies. DNAqua-net is composed of five working groups that contribute to these overarching goals: DNA Barcode References, Biotic Indices and Metrics, Field and Lab Protocols, Data Analysis and Storage, Implementation Strategy and Legal Issues (Leese *et al.*, 2016). Outputs from the DNAqua-net working groups are emerging (http://dnaqua.net/publications/) that will guide standardisation and improve molecular monitoring of European freshwater ecosystems. Within the UK, the UK DNA Working Group provides a forum for researchers and end users to identify priorities for the development of DNA-based monitoring tools. Collaborations between universities and end users (e.g. Natural England, Environment Agency, Scottish Environment Protection Agency) hold promise to procure and curate reference sequences for underrepresented taxa as well as identify optimal laboratory and bioinformatics workflows for different taxonomic groups. However, both researchers and end users must be prepared to contribute funding and investigator effort towards collection and storage of morphotaxonomically identified voucher specimens with DNA barcodes for different markers. This thesis has identified which UK vertebrates and invertebrates are not represented by public reference databases. It is now up to collaborative networks, such as the UK DNA Working Group, to use this information and generate reference sequences by partitioning the workload (e.g. taxonomic group, set number of species) between its members to benefit the metabarcoding community.

7.5 Future directions of eDNA monitoring in pond ecosystems

7.5.1 Biomonitoring

DNA metabarcoding and eDNA metabarcoding have been employed in biomonitoring programmes that use invertebrate communities to assess water quality and environmental stressors (Deiner *et al.*, 2017). For example, DNA metabarcoding has been implemented to assess the response of invertebrates to environmental stressors (e.g. dissolved oxygen, dissolved organic carbon, total nitrogen, conductivity, salinity, fine sediment, velocity) in streams (Emilson *et al.*, 2017; Beermann *et al.*, 2018; Macher *et al.*, 2018) and rivers (Li *et al.*, 2018d). DNA metabarcoding of invertebrates is close to being used for routine

biomonitoring by end users (Aylagas *et al.*, 2018; Hering *et al.*, 2018), whereas eDNA metabarcoding has some way to go due to the issues associated with this approach for invertebrates (see section 7.2). Nonetheless, eDNA metabarcoding could provide better inferences on the impact of stressors at catchment scale due to the different invertebrate taxa it identifies (Macher *et al.*, 2018). eDNA metabarcoding has also proven useful for biomonitoring when non-invertebrate taxa are considered. Andruszkiewicz *et al*. (2017) demonstrated the potential of eDNA metabarcoding for vertebrate-based biomonitoring schemes, where marine community composition differed according to depth. Kuzmina *et al*. (2018) also emphasised the potential of macrophyte communities inferred by eDNA metabarcoding to act as indicators of water quality and their biomonitoring potential.

7.5.2 Population genetics and distinguishing hybrids

As eDNA research continues to advance, there is growing interest in population genetics or diversity and the identification of individuals using eDNA analysis. Sigsgaard *et al*. (2016) were the first to use eDNA analysis to characterise genetic diversity of a whale shark (*Rhincodon typus*) population. They found more *R. typus* haplotypes in seawater than tissue samples. Scaling down to individuals, Wheat *et al*. (2016) used saliva left on partially consumed salmon carcasses to identify brown bear (*Ursos arctos*) individuals as an alternative to screening scat from bears. More recently, Parsons *et al*. (2018) used high-throughput sequencing of eDNA samples to unlock population structure of the elusive harbour porpoise (*Phocoena phocoena*). Elsewhere, eDNA analysis was used to identify non-native haplotypes of common carp (*Cyprinus carpio*), which may enable greater surveillance of invasion patterns as well as protection of native populations of aquatic species (Uchii, Doi & Minamoto, 2016). Ponds may allow eDNA researchers to fully explore the use of nuclear markers for population genetics. Their small size presents an opportunity to obtain meaningful eDNA samples from which to estimate allele frequencies and ground truth with conventional sampling. Collecting environmental samples from individual deposits (e.g. faeces, contacted vegetation and/or substrates) will allow genotyping wildlife that utilise ponds (e.g. amphibians, odonates, mammals), which may enable their dispersal to be inferred and provide greater understanding of population and habitat connectivity (De Meester *et al.*, 2005).

7.5.3 Disease management

Detection and management of disease in freshwater environments is crucial to preventing spread and further infection. Crayfish plague (*Aphanomyces astaci* [Schikora, 1906]) and chytrid fungi (*Batrachochytrium dendrobatidis* [Longcore, Pessier & Nichols, 1999] and *B. salamandrivorans* [Martel *et al.*, 2013]) pose major threats to pond biodiversity. Chytrid fungi have decimated amphibian populations and contributed to global decline and extinction risk of species (Walker *et al.*, 2007; Mosher, Huyvaert & Bailey, 2018). Microscopy or molecular techniques were once used to detect zoosporangium in host individuals but swabs were required from the host's skin or mouth (Mosher *et al.*, 2018). eDNA analysis presented an alternative avenue of diagnosis: water is sampled and filtered, followed by detection of chytrid zoospores using qPCR (Kirshtein *et al.*, 2007; Walker *et al.*, 2007; Schmidt *et al.*, 2013; Mosher *et al.*, 2018). A similar procedure was developed to detect crayfish plague spores, carried by invasive North American crayfish but lethal to European crayfish species (Strand *et al.*, 2014), and has since been multiplexed to allow simultaneous qPCR detection of host, vector, and pathogen from eDNA (Robinson *et al.*, 2018). eDNA metabarcoding may be the next logical step to screen for multiple freshwater diseases that threaten biodiversity, or to monitor host, threatened species, and pathogens simultaneously. Microbiome research is another field that has been pivotal to understanding chytrid fungus resistance and immunity in amphibian species, and cure development. Obtaining microbiome data has been dependent on whole body or ventral swabbing, but eDNA metabarcoding of bacterial communities may be an option where tissue samples are not available.

7.5.4 Citizen science

Ponds are poorly monitored in comparison to other freshwater habitats despite their biodiversity potential. However, citizen science combined with eDNA analysis has the potential to revolutionise pond monitoring and provide much needed long-term baseline data (Biggs *et al.*, 2016). One of the first studies to realise this potential focused on *T. cristatus* (Biggs *et al.*, 2015). Using eDNA analysis, volunteers detected great crested newt in 91.3% ponds (N = 239) and achieved a detection rate comparable to professional ecologists. Now, eDNA-based citizen science monitoring is being used in the 'Great

Crested Newt Detectives' project in Scotland to find new sites for the species and educate schools about the native herpetofauna (Amphibian & Reptile Conservation, 2016). Volunteers have also been utilised in the removal and eradication of invasive pygmy mussel (*Xenostrobus securis*), with successful eradication confirmed by visual search and eDNA survey (Miralles *et al.*, 2016).

Widespread integration of eDNA in citizen science projects is prevented only by the cost, time, and expertise required to process samples in the laboratory. Furthermore, project managers must decide whether to disseminate results only to volunteers or provide data for them to analyse. Nevertheless, eventual integration when these barriers are overcome will bolster public engagement with biodiversity monitoring and provide opportunity for education alongside obtaining large-scale biodiversity records for multiple species (Deiner *et al.*, 2017). In the context of citizen science, use of eDNA metabarcoding could be most effective and educational at BioBlitz events (http://www.bnhc.org.uk/bioblitz/) whilst targeted eDNA analysis will be most effective where local or national campaigns for species are already in place. Nonetheless, it is vital to recognise the impact of seeing wildlife in public outreach and education, thus eDNA analysis should not become the only method of citizen science.

7.6 Conclusions

Across the chapters of this thesis, I have demonstrated that eDNA analysis is a versatile and powerful tool for rapid biodiversity monitoring of freshwater ponds, particularly in the UK. Development of this tool for species and biotic assemblages not explored in this thesis should continue alongside further investigation into its weaknesses and limitations. eDNA analysis and the conventional tools employed in this thesis emphasise the biodiversity that pond ecosystems host, both individually and combined. Therefore, pondscapes deserve to be recognised in freshwater research, scientific monitoring, and policy. eDNA analysis will help achieve this recognition by enabling non-invasive, cost-effective, and time-efficient monitoring of aquatic, semi-aquatic and terrestrial biodiversity associated with pondscapes. The data generated by eDNA analysis for single species or entire communities has endless ecological applications that have been tested or identified in this thesis. Consequently, eDNA analysis will contribute to our understanding of the status and value of ponds, overcome the challenges associated with

monitoring these ecosystems, and inform the conservation and management of pondscapes.

References

Abell, R., Thieme, M.L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., Coad, B., Mandrak, N., Balderas, S.C., Bussing, W., Stiassny, M.L.J., Skelton, P., Allen, G.R., Unmack, P., Naseka, A., Ng, R., Sindorf, N., Robertson, J., Armijo, E., Higgins, J.V., Heibel, T.J., Wikramanayake, E., Olson, D., López, H.L., Reis, R.E., Lundberg, J.G., Pérez, M.H.S. & Petry, P. (2008) Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. *Bioscience*, **58**, 403–414.

Agersnap, S., Larsen, W.B., Knudsen, S.W., Strand, D., Thomsen, P.F., Hesselsøe, M., Mortensen, P.B., Vrålstad, T. & Møller, P.R. (2017) Monitoring of noble, signal and narrow-clawed crayfish using environmental DNA from freshwater samples. *PLoS ONE*, **12**, e0179261.

Aho, J., & Holopainen, I.J. (2000) Batch spawning of crucian carp (*Carassius carassius* (L.)) in mono- and multispecies communities. *Annales Zoologici Fennici*, **37**, 101–111.

Ahumada, J.A., Hurtado, J. & Lizcano, D. (2013) Monitoring the status and trends of tropical forest terrestrial vertebrate communities from camera trap data: a tool for conservation. *PLoS ONE*, **8**, e73707.

Akaike, H. (1973) Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, **60**, 255–265.

Alderton, E., Sayer, C.D., Davies, R., Lambert, S.J. & Axmacher, J.C. (2017) Buried alive: Aquatic plants survive in “ghost ponds” under agricultural fields. *Biological Conservation*, **212**, 105–110.

Almeida, D., Rodolfo, N., Sayer, C.D. & Copp, G.H. (2013) Seasonal use of ponds as foraging habitat by Eurasian otter with description of an alternative handling technique for common toad predation. *Folia Zoologica*, **62**, 214–221.

Andersen, K., Bird, K.L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjær, K.H., Orlando, L., Gilbert, M.T.P. & Willerslev, E. (2012) Meta-barcoding of “dirt” DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, **21**, 1966–1979.

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

Andújar, C., Arribas, P., Gray, C., Bruce, C., Woodward, G., Yu, D.W. & Vogler, A.P. (2017) Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. *Molecular Ecology*, **27**, 146–166.

ARG-UK. (2010). *ARG UK Advice Note 5: Great Crested Newt Habitat Suitability Index*. Amphibian and Reptile Groups of the United Kingdom.

Aylagas, E., Borja, Á., Muxika, I. & Rodríguez-Ezpeleta, N. (2018) Adapting metabarcoding-based benthic biomonitoring into routine marine ecological status assessment networks. *Ecological Indicators*, **95**, 194–202.

Aylward, M.L., Johnson, S.E., Sullivan, A.P., Perry, G.H. & Louis, E.E. (2018) A novel environmental DNA (eDNA) sampling method for aye-ayes from their feeding traces. *bioRxiv*, 272153.

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**, 2039–2044.

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

Baker, S.J., Niemiller, M.L., Stites, A.J., Ash, K.T., Davis, M.A., Dreslik, M.J. & Phillips, C.A. (2018b) Evaluation of environmental DNA to detect *Sistrurus catenatus* and *Ophidiomyces ophiodiicola* in crayfish burrows. *Conservation Genetics Resources*, 1–3.

Balian, E.V., Segers, H., Lévèque, C. & Martens, K. (2008) The Freshwater Animal Diversity Assessment: an overview of the results. *Hydrobiologia*, **595**, 627–637.

Bálint, M., Nowak, C., Márton, O., Pauls, S.U., Wittwer, C., Aramayo, J.L., Schulze, A., Chambert, T., Cocchiararo, B. & Jansen, M. (2018) Accuracy, limitations and cost efficiency of eDNA-based community survey in tropical frogs. *Molecular Ecology Resources*, **18**, 1415–1426.

Barnes, M.A., Turner, C.R., Jerde, C.L., Renshaw, M.A., Chadderton, W.L. & Lodge, D.M. (2014) Environmental Conditions Influence eDNA Persistence in Aquatic Systems. *Environmental Science & Technology*, **48**, 1819−1827.

Barnes, M.A. & Turner, C.R. (2015) The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, **17**, 1–17.

Bartout, P., Touchart, L., Terasmaa, J., Choffel, Q., Marzecova, A., Koff, T., Kapanen, G., Qsair, Z., Maleval, V., Millot, C., Saudubray, J. & Aldomany, M. (2015) A new approach to inventorying bodies of water, from local to global scale. *DIE ERDE*, **146**, 245–258.

Baselga, A. & Orme, C.D.L. (2012) betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution*, **3**, 808–812.

Bass, J. (1998) *Last-instar larvae and pupae of the Simuliidae of Britain and Ireland: A key with brief ecological notes*. Freshwater Biological Association, Ambleside.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**, 1–48.

Battersby, J.E. & Greenwood, J.J.D. (2004) Monitoring terrestrial mammals in the UK: past, present and future, using lessons from the bird world. *Mammal Review*, **34**, 3–29.

Battersby, J. & Tracking Mammals Partnership (2005) *UK Mammals: Species Status and Population Trends. First Report by the Tracking Mammals Partnership*. Joint Nature Conservation Committee /Tracking Mammals Partnership, Peterborough.

Batzer, D.P., Pusateri, C.R. & Vetter, R. (2000) IMPACTS OF FISH PREDATION ON MARSH INVERTEBRATES: DIRECT AND INDIRECT EFFECTS. *Wetlands*, **20**, 307–312.

Beebee, T.J.C. (1997) Changes in dewpond numbers and amphibian diversity over 20 years on chalk downland in Sussex, England. *Biological Conservation*, **81**, 215–219.

Beermann, A.J., Zizka, V.M.A., Elbrecht, V., Baranov, V. & Leese, F. (2018) DNA metabarcoding reveals the complex and hidden responses of chironomids to multiple stressors. *Environmental Sciences Europe*, **30**, 26.

Bellemain, E., Patricio, H., Gray, T., Guegan, F., Valentini, A., Miaud, C. & Dejean, T. (2016) Trails of river monsters: Detecting critically endangered Mekong giant catfish *Pangasianodon gigas* using environmental DNA. *Global Ecology and Conservation*, **7**, 148–156.

Bendell, B.E. & McNicol, D.K. (1995) Lake acidity, fish predation, and the distribution and abundance of some littoral insects. *Hydrobiologia*, **302**, 133–145.

Beutel, M.W. & Larson, L. (2015) Pathogen removal from urban pond outflow using rock biofilters. *Ecological Engineering*, **78**, 72–78.

Biggs, J., Williams, P.J., Corfield, A., Whitfield, M.A., Barr, C.J. & Cummins, C.P. (1996) *Pond Survey 1996: stage 1 scoping study*.

Biggs, J., Fox, G. & Nicolet, P. (1998) *A guide to the methods of the National Pond Survey*. Pond Action, Oxford.

Biggs, J., Williams, P., Whitfield, M., Nicolet, P. & Weatherby, A. (2005) 15 years of pond assessment in Britain: results and lessons learned from the work of Pond Conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **15**, 693–714.

Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C. & Griffiths, R.A. (2014) *Analytical and methodological development for improved surveillance of the Great Crested Newt*. Defra Project WC1067.

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.

Biggs, J., von Fumetti, S. & Kelly-Quinn, M. (2016) The importance of small waterbodies for biodiversity and ecosystem services: implications for policy makers. *Hydrobiologia*, **793**, 3–39.

Bista, I., Carvalho, G., Walsh, K., Christmas, M., Hajibabaei, M., Kille, P., Lallias, D. & Creer, S. (2015) Monitoring lake ecosystem health using metabarcoding of environmental DNA: temporal persistence and ecological relevance. *Genome*, **58**, 197.

Bista, I., Carvalho, G.R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., Christmas, M. & Creer, S. (2017) Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, **8**, 14087.

Bjørnstad, O.N. (2017) ncf: Spatial Covariance Functions. R package version 1.1-3.

Blackman, R.C., Constable, D., Hahn, C., Sheard, A.M., Durkota, J., Hänfling, B. & Lawson Handley, L. (2017) Detection of a new non-native freshwater species by DNA metabarcoding of environmental samples – first record of *Gammarus fossarum* in the UK. *Aquatic Invasions*, **12**, 177-189.

Blackman, R.C., Benucci, M., Donnelly, R., Hänfling, B., Harper, L.R., Kimbell, H., Sellers, G.S., Sheard, A.M., Watson, H.V. & Lawson-Handley, L. (2018) Targeting the Invaders – Targeted Detection of Four Priority Freshwater Invasive Non-Native Species Using Environmental DNA. *PeerJ Preprints*.

Boessenkool, S., Epp, L.S., Haile, J., Bellemain, E., Edwards, M., Coissac, E., Willerslev, E. & Brochmann, C. (2012) Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Molecular Ecology*, **21**, 1806–1815.

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

Boix D., Biggs J., Céréghino R., Hull A.P., Kalettka T. & Oertli B. (2012) Pond research and management in Europe: "Small is Beautiful". *Hydrobiologia*, **689**, 1–9.

Bolger, A. M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bonesi, L. & Macdonald, D.W. (2004) Differential habitat use promotes sustainable coexistence between the specialist otter and the generalist mink. *Oikos*, **106**, 509–519.

Boothby J. (1997) Pond conservation: towards a delineation of pondscape. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **7**, 127–132.

Boothby J. & Hull A.P. (1997) A census of ponds in Cheshire, North West England. *Aquatic Conservation: Marine and Freshwater Ecosystems* **7**, 75–79.

Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the Spatial Component of Ecological Variation. *Ecology*, **73**, 1045–1055.

Brandon-Mong, G.-J., Gan, H.-M., Sing, K.-W., Lee, P.-S., Lim, P.-E. & Wilson, J.-J. (2015) DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bulletin of Entomological Research*, **105**, 717–727.

Brautigam A. (1999) The freshwater biodiversity crisis. *World Conservation*, **2**, 4–5.

Briers, R.A. & Biggs, J. (2003) Indicator taxa for the conservation of pond invertebrate diversity. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **13**, 323–330.

Brönmark, C. & Hansson, L.-A. (2002) Environmental issues in lakes and ponds: current state and perspectives. *Environmental Conservation*, **29**, 290–307.

Bronner, I.F., Quail, M.A., Turner, D.J. & Swerdlow, H. (2009) Improved Protocols for Illumina Sequencing. *Current Protocols in Human Genetics*, **18**, 18.2.1–18.2.42.

Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Machler, M. & Bolker, B.M. (2017) glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, **9**, 378–400.

Brooks, S. & Cham, S. (2014) *Field Guide to the Dragonflies & Damselflies of Great Britain and Ireland*. Bloomsbury Publishing Plc, London.

Burton, A.C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J.T., Bayne, E. & Boutin, S. (2015) REVIEW: Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, **52**, 675–685.

Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.-F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Hernández Morcillo, M., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vié, J.-C. & Watson, R. (2010) Global biodiversity: indicators of recent declines. *Science*, **328**, 1164–1168.

Buxton, A.S., Groombridge, J.J. & Griffiths, R.A. (2017a) Is the detection of aquatic environmental DNA influenced by substrate type? *PLoS ONE*, **12**, e0183371.

Buxton, A.S., Groombridge, J.J., Zakaria, N.B. & Griffiths, R.A. (2017b) Seasonal variation in environmental DNA in relation to population size and environmental factors. *Scientific Reports*, **7**, 46294.

Buxton, A.S., Groombridge, J.J. & Griffiths, R.A. (2018) Seasonal variation in environmental DNA detection in sediment and water samples. *PLoS ONE*, **13**, e0191737.

Bylemans, J., Furlan, E.M., Hardy, C.M., McGuffie, P., Lintermans, M. & Gleeson, D.M. (2017) An environmental DNA-based method for monitoring spawning activity: a case study, using the endangered Macquarie perch (*Macquaria australasica*). *Methods in Ecology and Evolution*, **8**, 646–655.

Bylemans, J., Gleeson, D.M., Hardy, C.M. & Furlan, E. (2018) Toward an ecoregion scale evaluation of eDNA metabarcoding primers: A case study for the freshwater fish biodiversity of the Murray-Darling Basin (Australia). *Ecology and Evolution*, **8**, 8697–8712.

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

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**, 22908.

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

Carew, M.E., Coleman, R.A. & Hoffmann, A.A. (2018a) Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding? *PeerJ*, **6**, e4980.

Carew, M.E., Kellar, C.R., Pettigrove, V.J. & Hoffmann, A.A. (2018b) Can high-throughput sequencing detect macroinvertebrate diversity for routine monitoring of an urban river? *Ecological Indicators*, **85**, 440–450.

Céréghino, R., Biggs, J., Oertli, B. & Declerck, S. (2008) The ecology of European ponds: defining the characteristics of a neglected freshwater habitat. *Hydrobiologia*, **597**, 1–6.

Chan, K. (2011) *Can great crested newts (Triturus cristatus) coexist with fish?* Masters Thesis, University College London (UCL).

Cilleros, K., Valentini, A., Allard, L., Dejean, T., Etienne, R., Grenouillet, G., Iribar, A., Taberlet, P., Vigouroux, R. & Brosse, S. (2019) Unlocking biodiversity and conservation studies in high diversity environments using environmental DNA (eDNA): a test with Guianese freshwater fishes. *Molecular Ecology Resources*, **19**, 27–46.

Civade, R., Dejean, T., Valentini, A., Roset, N., Raymond, J.-C., Bonin, A., Taberlet, P. & Pont, D. (2016) Spatial Representativeness of Environmental DNA Metabarcoding Signal for Fish Biodiversity Assessment in a Natural Freshwater System. *PLoS ONE*, **11**, e0157366.

Clare, E.L., Chain, F.J.J., Littlefair, J.E. & Cristescu, M.E. (2016) The effects of parameter choice on defining molecular operational taxonomic units and resulting ecological analyses of metabarcoding data. *Genome*, **59**, 981–990.

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

Cohen, J. (1960) A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, **20**, 37–46.

Coissac, E., Riaz, T. & Puillandre, N. (2012) Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology*, **21**, 1834–1847.

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

Comtet, T., Sandionigi, A., Viard, F. & Casiraghi, M. (2015) DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions*, **17**, 905–922.

Copp, G.H., Černý, J. & Kováč, V. (2008a) Growth and morphology of an endangered native freshwater fish, crucian carp *Carassius carassius*, in an English ornamental pond. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **18**, 32–43.

Copp, G.H., Warrington, S. & Wesley, K.J. (2008b) Management of an ornamental pond as a conservation site for a threatened native fish species, crucian carp *Carassius carassius*. *Hydrobiologia*, **597**, 149-155.

Copp, G.H. & Sayer, C.D. (2010) *Norfolk Biodiversity Action Plan–Local Species Action Plan for Crucian Carp (Carassius carassius)*. Norfolk Biodiversity Partnership Reference: LS/ 3. Centre for Environment, Fisheries & Aquaculture Science, Lowestoft, Suffolk.

Cowart, D.A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J. & Arnaud-Haond, S. (2015) Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. *PLoS ONE*, **10**, e0117562.

Craine, J., Cannon, M., Elmore, A., Guinn, S. & Fierer, N. (2017) DNA metabarcoding potentially reveals multi-assemblage eutrophication responses in an eastern North American river. *bioRxiv*, 186452.

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 & Evolution*, **29**, 566–571.

Curry, C.J., Gibson, J.F., Shokralla, S., Hajibabaei, M. & Baird, D.J. (2018) Identifying North American freshwater invertebrates using DNA barcodes: are existing COI sequence libraries fit for purpose? *Freshwater Science*, **37**, 178–189.

Cusack, J.J., Dickman, A.J., Rowcliffe, J.M., Carbone, C., Macdonald, D.W. & Coulson, T. (2015) Random versus Game Trail-Based Camera Trap Placement Strategy for Monitoring Terrestrial Mammal Communities. *PLoS ONE*, **10**, e0126373.

Davies, B., Biggs, J., Williams, P., Whitfield, M., Nicolet, P., Sear, D., Bray, S. & Maund, S. (2008) Comparative biodiversity of aquatic habitats in the European agricultural landscape. *Agriculture, Ecosystems & Environment*, **125**, 1–8.

Davison, P.I., Créach, V., Liang, W.-J., Andreou, D., Britton, J.R. & Copp, G.H. (2016) Laboratory and field validation of a simple method for detecting four species of non-native freshwater fish using eDNA. *Journal of Fish Biology*, **89**, 1782–1793.

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.

De Meester, L., Declerck, S., Stoks, R., Louette, G., Van De Meutter, F., De Bie, T., Michels, E. & Brendonck, L. (2005) Ponds and pools as model systems in conservation biology, ecology and evolutionary biology. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **15**, 715–725.

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.

Deiner, K., Fronhofer, E.A., Mächler, E., Walser, J.-C. & Altermatt, F. (2016) Environmental DNA reveals that rivers are conveyer belts of biodiversity information. *Nature Communications*, **7**, 12544.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., de Vere, N., Pfrender, M.E. & Bernatchez, L. (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, **26**, 5872–5895.

Deiner, K., Lopez, J., Bourne, S., Holman, L., Seymour, M., Grey, E.K., Lacoursière, A., Li, Y., Renshaw, M.A., Pfrender, M.E., Rius, M., Bernatchez, L. & Lodge, D.M. (2018) Optimising the detection of marine taxonomic richness using environmental DNA metabarcoding: the effects of filter material, pore size and extraction method. *Metabarcoding and Metagenomics*, **2**, e28963.

Denoël, M. & Ficetola, G.F. (2008) Conservation of newt guilds in an agricultural landscape of Belgium: the importance of aquatic and terrestrial habitats. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **18**, 714–728.

Denoël, M., Perez, A., Cornet, Y. & Ficetola, G.F. (2013) Similar local and landscape processes affect both a common and a rare newt species. *PLoS ONE*, **8**, e62727.

Djurhuus, A., Port, J., Closek, C., Yamahara, K., Romero-Maraccini, O., Walz, K., Goldsmith, D., Michisaki, R., Boehm, A., Breitbart, M. & Chavez, F. (2017) Evaluation of filtration and DNA extraction methods for environmental DNA biodiversity assessments across multiple trophic levels. *Frontiers in Marine Science*, **4**, 314.

Dobson, M., Pawley, S., Fletcher, M. & Powell, A. (2012) *Guide to Freshwater Invertebrates*. Freshwater Biological Association, Ambleside.

Doi, H., Katano, I., Sakata, Y., Souma, R., Kosuge, T., Nagano, M., Ikeda, K., Yano, K. & Tojo, K. (2017) Detection of an endangered aquatic heteropteran using environmental DNA in a wetland ecosystem. *Royal Society Open Science*, **4**, 170568.

Dorazio, R.M. & Erickson, R.A. (2017) eDNAoccupancy: An R package for multi-scale occupancy modeling of environmental DNA data. *Molecular Ecology Resources*, **18**, 368-380.

Downing, J.A., Prairie, Y.T., Cole, J.J., Duarte, C.M., Tranvik, L.J., Striegl, R.G., McDowell, W.H., Kortelainen, P., Caraco, N.F., Melack, J.M. & Middelburg, J.J. (2006) The global abundance and size distribution of lakes, ponds, and impoundments. *Limnology and Oceanography*, **51**, 2388–2397.

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.J. & Sullivan, C.A. (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews*, **81**, 163–182.

Dunn, N., Priestley, V., Herraiz, A., Arnold, R. & Savolainen, V. (2017) Behavior and season affect crayfish detection and density inference using environmental DNA. *Ecology and Evolution*, **7**, 7777–7785.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Edington, J.M. & Hildrew, A.G. (1995) *A revised key to the caseless caddis larvae of the British Isles with notes on their ecology*. Freshwater Biological Association, Ambleside.

Egeter, B., Peixoto, S., Brito, J.C., Jarman, S., Puppo, P. & Velo-Antón, G. (2018) Challenges for assessing vertebrate diversity in turbid Saharan water-bodies using environmental DNA. *Genome*, **61**, 807–814.

Eichmiller, J.J., Bajer, P.G. & Sorensen, P.W. (2014) The relationship between the distribution of common carp and their environmental DNA in a small lake. *PLoS ONE*, **9**, e112611.

Eichmiller, J., Best, S.E. & Sorensen, P.W. (2016) Effects of temperature and trophic state on degradation of environmental DNA in lake water. *Environmental Science & Technology*, **50**, 1859-1867.

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

Elbrecht, V. & Leese, F. (2017) Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science*, **5**.

Elbrecht, V., Peinert, B. & Leese, F. (2017a) Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecology and Evolution*, **7**, 6918–6926.

Elbrecht, V., Vamos, E.E., Meissner, K., Aroviita, J. & Leese, F. (2017b) Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, **8**, 1265–1275.

Elbrecht, V., Hebert, P.D.N. & Steinke, D. (2018) Slippage of degenerate primers can cause variation in amplicon length. *Scientific Reports*, **8**, 10999.

Elliott, J.M. (2009) *Freshwater Megaloptera and Neuroptera of Britain and Ireland: keys to adults and larvae, and a review of their ecology*. Freshwater Biological Association, Ambleside.

Elliott, J.M. & Humpesch, U.H. (2010) *Mayfly larvae (Ephemeroptera) of Britain and Ireland: keys and a review of their ecology*. Freshwater Biological Association, Ambleside.

Elliott, J.M. & Dobson, M. (2015) *Freshwater leeches of Britain and Ireland: keys to the Hirudinea and a review of their ecology*. Freshwater Biological Association, Ambleside.

Emilson, C.E., Thompson, D.G., Venier, L.A., Porter, T.M., Swystun, T., Chartrand, D., Capell, S. & Hajibabaei, M. (2017) DNA metabarcoding and morphological macroinvertebrate metrics reveal the same changes in boreal watersheds across an environmental gradient. *Scientific Reports*, **7**, 12777.

Emson, D., Sayer, C.D., Bennion, H., Patmore, I.R. & Rioual, P. (2017) Mission possible: diatoms can be used to infer past duckweed (lemnoid Araceae) dominance in ponds. *Journal of Paleolimnology*, **60**, 209–221.

Environment Agency (2003) *Crucian carp field guide*. National Coarse Fish Centre, Environment Agency, Bristol.

European Commission (2000) Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy. Accessed on 1st December 2018. https://publications.europa.eu/en/publication-detail/-/publication/70e52c10-85a1-4e97-8218-ed56d597ed05/language-en

Evans, D.M., Kitson, J.J.N., Lunt, D.H., Straw, N.A. & Pocock, M.J.O. (2016a) Merging DNA metabarcoding and ecological network analysis to understand and build resilient terrestrial ecosystems. *Functional Ecology*, **30**, 1904–1916.

Evans, N.T., Olds, B.P., Renshaw, M.A., Turner, C.R., Li, Y., Jerde, C.L., Mahon, A.R., Pfrender, M.E., Lamberti, G.A. & Lodge, D.M. (2016b) Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*, **16**, 29–41.

Evans, N.T., Li, Y., Renshaw, M.A., Olds, B.P., Deiner, K., Turner, C.R., Jerde, C.L., Lodge, D.M., Lamberti, G.A. & Pfrender, M.E. (2017a) Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, **74**, 1362–1374.

Evans, N.T., Shirey, P.D., Wieringa, J.G., Mahon, A.R. & Lamberti, G.A. (2017b) Comparative Cost and Effort of Fish Distribution Detection via Environmental DNA Analysis and Electrofishing. *Fisheries*, **42**, 90–99.

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.

Ficetola, G.F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P. & Pompanon, F. (2010) An *In silico* approach for the evaluation of DNA barcodes. *BMC Genomics*, **11**, 434.

Ficetola, G.F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Lopes, C.M., Boyer, F., Pompanon, F., Rayé, G. & Taberlet, P. (2015) Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, **15**, 543–556.

Ficetola, G.F., Taberlet, P. & Coissac, E. (2016) How to limit false positives in environmental DNA and metabarcoding? *Molecular Ecology Resources*, **16**, 604–607.

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.

Fox, J. & Weisberg S. (2011) *An R Companion to Applied Regression, Second Edition*. Sage, Thousand Oaks, CA.

Freshwater Habitats Trust (2015) *Pond Habitat Survey: Survey Manual*. Freshwater Habitats Trust, Oxford.

Friday, L.E. (1988) A key to the adults of British water beetles. *Field Studies*, **7**, 1–151.

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.

Gantz, C.A., Renshaw, M.A., Erickson, D., Lodge, D.M. & Egan, S.P. (2018) Environmental DNA detection of aquatic invasive plants in lab mesocosm and natural field conditions. *Biological Invasions*, **20**, 2535–2552.

Gaughran, A., Kelly, D.J., MacWhite, T., Mullen, E., Maher, P., Good, M. & Marples, N.M. (2018) Super-ranging. A new ranging strategy in European badgers. *PLoS ONE*, **13**, e0191818.

Gee, J.H.R., Smith, B.D., Lee, K.M. & Griffiths, S.W. (1997) The ecological basis of freshwater pond management for biodiversity. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **7**, 91–104.

Geerts, A.N., Boets, P., Van den Heede S., Goethals, P. & Van der heyden, C. (2018) A search for standardized protocols to detect alien invasive crayfish based on environmental DNA (eDNA): A lab and field evaluation. *Ecological Indicators*, **84**, 564–572.

Geist, J. (2011) Integrative freshwater ecology and biodiversity conservation. *Ecological Indicators*, **11**, 1507–1516.

Geller, J., Meyer, C., Parker, M. & Hawk, H. (2013) Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, **13**, 851–861.

Gledhill, T., Sutcliffe, D.W. & Williams, W.D. (1993) *British freshwater Crustacea Malacostraca: a key with ecological notes*. Freshwater Biological Association, Ambleside.

Gledhill, D.G., James, P. & Davies, D.H. (2008) Pond density as a determinant of aquatic species richness in an urban landscape. *Landscape Ecology*, **23**, 1219–1230.

Glen, A.S., Cockburn, S., Nichols, M., Ekanayake, J. & Warburton, B. (2013) Optimising camera traps for monitoring small mammals. *PLoS ONE*, **8**, e67940.

Goertzen, D. & Suhling, F. (2012) Promoting dragonfly diversity in cities: major determinants and implications for urban pond design. *Journal of Insect Conservation*, **17**, 399–409.

Goldberg, C.S., Turner, C.R., Deiner, K., Klymus, K.E., Thomsen, P.F., Murphy, M.A., Spear, S.F., McKee, A., Oyler-McCance, S.J., Cornman, R.S., Laramie, M.B., Mahon, A.R., Lance, R.F., Pilliod, D.S., Strickler, K.M., Waits, L.P., Fremier, A.K., Takahara, T., Herder, J.E. & Taberlet, P. (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, **7**, 1299–1307.

Goldberg, C.S., Strickler, K.M. & 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**, 695–703.

Gounand, I., Harvey, E., Little, C.J. & Altermatt, F. (2018) Meta-Ecosystems 2.0: Rooting the Theory into the Field. *Trends in Ecology & Evolution*, **33**, 36–46.

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., Pfrender, M.E., Prowse, T.A.A., Renshaw, M.A. & Lodge, D.M. (2018) Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Scientific Reports*, **8**, 8843.

Griffith, D., Veech, J. & Marsh, C. (2016) cooccur: Probabilistic Species Co-Occurrence Analysis in R. *Journal of Statistical Software*, **69**, 1–17.

Gustafson, D.H., Pettersson, C.J. & Malmgren, J.C. (2006) Great crested newts (*Triturus cristatus*) as indicators of aquatic plant diversity. *The Herpetological Journal*, **16**, 347–352.

Gustafson, D.H., Malmgren, J.C. & Mikusiński, G. (2011) Terrestrial Habitat Predicts use of Aquatic Habitat for Breeding Purposes — A Study on the Great Crested Newt (*Triturus cristatus*). *Annales Zoologici Fennici*, **48**, 295–307.

Haas, K., Köhler, U., Diehl, S., Köhler, P., Dietrich, S., Holler, S., Jaensch, A., Niedermaier, M. & Vilsmeier, J. (2007) Influence of fish on habitat choice of water birds: a whole system experiment. *Ecology*, **88**, 2915–2925.

Haase, P., Pauls, S.U., Schindehütte, K. & Sundermann, A. (2010) First audit of macroinvertebrate samples from an EU Water Framework Directive monitoring program: human error greatly lowers precision of assessment results. *Journal of the North American Benthological Society*, **29**, 1279–1291.

Hajibabaei, M., Baird, D.J., Fahner, N.A., Beiko, R. & Golding, G.B. (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*, **371**, 20150330.

Hänfling, B., Bolton, P., Harley, M. & Carvalho, G.R. (2005) A molecular approach to detect hybridisation between crucian carp (*Carassius carassius*) and non-indigenous carp species (*Carassius* spp. and *Cyprinus carpio*). *Freshwater Biology*, **50**, 403–417.

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.

Hänfling, B., Lawson Handley, L., Lunt, D., Shum, P., Winfield, I. & Read, D.S. (2017) *A Review of Recent Advances in Genetic Methods to Identify Improvements in CAMERAS Partners Monitoring Activities*. University of Hull, Hull, UK.

Hardie, S.A., Barmuta, L.A. & White, R.W.G. (2006) Comparison of Day and Night Fyke Netting, Electrofishing and Snorkelling for Monitoring a Population of the Threatened Golden Galaxias (*Galaxias auratus*). *Hydrobiologia*, **560**, 145–158.

Harper, L.R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H.C., Gough, K.C., Lewis, E., Adams, I.P., Brotherton, P., Phillips, S. & Hänfling, B. (2018a) Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*). *Ecology and Evolution*, **8**, 6330–6441.

Harper, L.R., Handley, L.L., Hahn, C., Boonham, N., Rees, H.C., Lewis, E., Adams, I.P., Brotherton, P., Phillips, S. & Hänfling, B. (2018b) Ground-truthing environmental DNA (eDNA) metabarcoding for ecological hypothesis testing at the pondscape. *bioRxiv*, 278309.

Harper, K., Anucha, P., Turnbull, J., Bean, C. & Leaver, M. (2018c) Searching for a signal: Environmental DNA (eDNA) for the detection of invasive signal crayfish, *Pacifastacus leniusculus* (Dana, 1852). *Management of Biological Invasions*, **9**, 137–148.

Harper, L.R., Buxton, A.S., Rees, H.C., Bruce, K., Brys, R., Halfmaerten, D., Read, D.S., Watson, H.V., Sayer, C.D., Jones, E.P., Priestley, V., Mächler, E., Múrria, C., Garcés-Pastor, S., Medupin, C., Burgess, K., Benson, G., Boonham, N., Griffiths, R.A., Lawson Handley, L. & Hänfling, B. (2019a) Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. *Hydrobiologia*, **826**, 25–41.

Harper, L.R., Griffiths, N.P., Lawson Handley, L., Sayer, C.D., Read, D.S., Harper, K.J., Blackman, R.C., Li, J. & Hänfling, B. (2019b) Development and application of environmental DNA surveillance for the threatened crucian carp (*Carassius carassius*). *Freshwater Biology*, **64**, 93–107.

Harris, S. & Yalden, D.W. (2004) An integrated monitoring programme for terrestrial mammals in Britain. *Mammal Review*, **34**, 157–167.

Hartel, T., Nemes, S. & Oellerer, K. (2010) Using connectivity metrics and niche modelling to explore the occurrence of the northern crested newt *Triturus cristatus* (Amphibia, Caudata) in a traditionally managed landscape. *The Environmentalist*, **37**, 195–200.

Hassall, C., Hollinshead, J. & Hull, A. (2011) Environmental correlates of plant and invertebrate species richness in ponds. *Biodiversity and Conservation*, **20**, 3189–3222.

Hassall, C., Hollinshead, J. & Hull, A. (2012) Temporal dynamics of aquatic communities and implications for pond conservation. *Biodiversity and Conservation*, **21**, 829–852.

Heath, D.J. & Whitehead, A. (1992) A survey of pond loss in Essex, South-east England. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **2**, 267–273.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B*, **270**, 313–321.

Hering, D., Borja, A., Jones, J.I., Pont, D., Boets, P., Bouchez, A., Bruce, K., Drakare, S., Hänfling, B., Kahlert, M., Leese, F., Meissner, K., Mergen, P., Reyjol, Y., Segurado, P., Vogler, A. & Kelly, M. (2018) Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. *Water Research*, **138**, 192–205.

Hervé, M. (2015) RVAideMemoire: Diverse basic statistical and graphical functions. R package version 0.9-36.

Hill, M.J., Sayer, C.D. & Wood, P.J. (2016) When is the best time to sample aquatic macroinvertebrates in ponds for biodiversity assessment? *Environmental Monitoring and Assessment*, **188**, 194.

Hill, M.J., Hassall, C., Oertli, B., Fahrig, L., Robson, B.J., Biggs, J., Samways, M.J., Usio, N., Takamura, N., Krishnaswamy, J. & Wood, P.J. (2018) New policy directions for global pond conservation. *Conservation Letters*, **142**, e12447.

Hinlo, R., Furlan, E., Suitor, L. & Gleeson, D. (2017a) Environmental DNA monitoring and management of invasive fish: comparison of eDNA and fyke netting. *Management of Biological Invasions*, **8**, 89–100.

Hinlo, R., Gleeson, D., Lintermans, M. & Furlan, E. (2017b) Methods to maximise recovery of environmental DNA from water samples. *PLoS ONE*, **12**, e0179251.

Hosmer, D.W. & Lemeshow, S. (2000) Multiple Logistic Regression. In: *Applied Logistic Regression*. John Wiley & Sons, Inc., pp. 31–46.

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.

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

Illumina (2011) *Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System*. Illumina Technical Note.

Ishige, T., Miya, M., Ushio, M., Sado, T., Ushioda, M., Maebashi, K., Yonechi, R., Lagan, P. & Matsubayashi, H. (2017) Tropical-forest mammals as detected by environmental DNA at natural saltlicks in Borneo. *Biological Conservation*, **210**, 281–285.

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.

Jarvis, L.E. (2010) Non-consumptive effects of predatory three-spined sticklebacks (*Gasterosteus aculeatus*) on great crested newt (*Triturus cristatus*) embryos. *The Herpetological Journal*, **20**, 271–275.

Jarvis, L.E. (2012) *Microhabitat preferences of the great crested newt Triturus cristatus in a woodland area*. PhD Thesis, The Open University.

Jeffries, D.L., Copp, G.H., Lawson Handley, L., Olsén, K.H., Sayer, C.D. & Hänfling, B. (2016) Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*, L. *Molecular Ecology*, **25**, 2997-3018.

Jeffries, D.L., Copp, G.H., Maes, G.E., Lawson Handley, L., Sayer, C.D. & Hänfling, B. (2017) Genetic evidence challenges the native status of a threatened freshwater fish (*Carassius carassius*) in England. *Ecology and Evolution*, **7**, 2871-2882.

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.

Joint Nature Conservation Committee & Defra, (2012) UK post-2010 biodiversity framework. Joint Nature Conservation Committee, Peterborough. Accessed on 1st December 2018. http://jncc.defra.gov.uk/page-6189

Joint Nature Conservation Committee (2018) UK BAP priority terrestrial mammal species. Accessed on 1st December 2018. http://jncc.defra.gov.uk/page-5170

Joly, P., Miaud, C., Lehmann, A. & Grolet, O. (2001) Habitat Matrix Effects on Pond Occupancy in Newts. *Conservation Biology*, **15**, 239–248.

Kazanjian, G., Flury, S., Attermeyer, K., Kalettka, T., Kleeberg, A., Premke, K., Köhler, J. & Hilt, S. (2018) Primary production in nutrient-rich kettle holes and consequences for nutrient and carbon cycling. *Hydrobiologia*, **806**, 77–93.

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

Kelly, R.P., O'Donnell, J.L., Lowell, N.C., Shelton, A.O., Samhouri, J.F., Hennessey, S.M., Feist, B.E. & Williams, G.D. (2016) Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ*, **4**, e2444.

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

Kirshtein, J.D., Anderson, C.W., Wood, J.S., Longcore, J.E. & Voytek, M.A. (2007) Quantitative PCR detection of *Batrachochytrium dendrobatidis* DNA from sediments and water. *Diseases of Aquatic Organisms*, **77**, 11–15.

Kitson, J.J.N., Hahn, C., Sands, R.J., Straw, N.A., Evans, D.M. & Lunt, D.H. (2019) Detecting host-parasitoid interactions in an invasive Lepidopteran using nested tagging DNA-metabarcoding. *Molecular Ecology*, **28**, 471–483.

Klymus, K.E., Richter, C.A., Chapman, D.C. & Paukert, C. (2015) Quantification of eDNA shedding rates from invasive bighead carp *Hypophthalmichthys nobilis* and silver carp *Hypophthalmichthys molitrix*. *Biological Conservation*, **183**, 77–84.

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

Klymus, K.E., Richter, C.A., Thompson, N. & 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**, 54.

Kucherenko, A., Herman, J.E., Iii, E.M.E. & Urakawa, H. (2018) Terrestrial Snake Environmental DNA Accumulation and Degradation Dynamics and its Environmental Application. *Herpetologica*, **74**, 38–49.

Kundu, S., Kumar, V., Tyagi, K. & Chandra, K. (2018) Environmental DNA (eDNA) testing for detection of freshwater turtles in a temple pond. *Herpetology Notes*, **11**, 369–371.

Kuzmina, M.L., Braukmann, T.W.A. & Zakharov, E.V. (2018) Finding the pond through the weeds: eDNA reveals underestimated diversity of pondweeds. *Applications in Plant Sciences*, **6**, e01155.

Lacoursière-Roussel, A., Dubois, Y., Normandeau, E. & Bernatchez, L. (2016a) Improving herpetological surveys in eastern North America using the environmental DNA method. *Genome*, **59**, 991–1007.

Lacoursière-Roussel, A., Rosabal, M. & Bernatchez, L. (2016b) Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. *Molecular Ecology Resources*, **16**, 1401-1414.

Langton, T., Beckett, C. & Foster, J. (2001) *Great crested newt conservation handbook*. Froglife, Halesworth.

Larsson, A. (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, **30**, 3276–3278.

Lauridsen, T.L., Jeppesen, E. & Andersen, F.Ø. (1993) Colonization of submerged macrophytes in shallow fish manipulated Lake Væng: impact of sediment composition and waterfowl grazing. *Aquatic Botany*, **46**, 1–15.

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

Lawson Handley, L.J., Read, D., Winfield, I., Kimbell, H., Johnson, H., Li, J., Hahn, C., Blackman, R., Wilcox, R., Donnelly, R., Szitenberg, A. & Hänfling, B. (2018) Temporal and spatial variation in distribution of fish environmental DNA in England's largest lake. *bioRxiv*, 376400.

Lee, P.-S., Sing, K.-W. & Wilson, J.-J. (2015) Reading Mammal Diversity from Flies: The Persistence Period of Amplifiable Mammal mtDNA in Blowfly Guts (*Chrysomya megacephala*) and a New DNA Mini-Barcode Target. *PLoS ONE*, **10**, e0123871.

Leese, F., Altermatt, F., Bouchez, A., Ekrem, T., Hering, D., Meissner, K., Mergen, P., Pawlowski, J., Piggott, J., Rimet, F., Steinke, D., Taberlet, P., Weigand, A., Abarenkov, K., Beja, P., Bervoets, L., Björnsdóttir, S., Boets, P., Boggero, A., Bones, A., Borja, Á., Bruce, K., Bursić, V., Carlsson, J., Čiampor, F., Čiamporová-Zatovičová, Z., Coissac, E., Costa, F., Costache, M., Creer, S., Csabai, Z., Deiner, K., DelValls, Á., Drakare, S., Duarte, S., Eleršek, T., Fazi, S., Fišer, C., Flot, J.-F., Fonseca, V., Fontaneto, D., Grabowski, M., Graf, W., Guðbrandsson, J., Hellström, M., Hershkovitz, Y., Hollingsworth, P., Japoshvili, B., Jones, J., Kahlert, M., Kalamujic Stroil, B., Kasapidis, P., Kelly, M., Kelly-Quinn, M., Keskin, E., Kõljalg, U., Ljubešić, Z., Maček, I., Mächler, E., Mahon, A., Marečková, M., Mejdandzic, M., Mircheva, G., Montagna, M., Moritz, C., Mulk, V., Naumoski, A., Navodaru, I., Padisák, J., Pálsson, S., Panksep, K., Penev, L., Petrusek, A., Pfannkuchen, M., Primmer, C., Rinkevich, B., Rotter, A., Schmidt-Kloiber, A., Segurado, P., Speksnijder, A., Stoev, P., Strand, M., Šulčius, S., Sundberg, P., Traugott, M., Tsigenopoulos, C., Turon, X., Valentini, A., van der Hoorn, B., Várbíró, G., Vasquez Hadjilyra, M., Viguri, J., Vitonytė, I., Vogler, A., Vrålstad, T., Wägele, W., Wenne, R., Winding, A., Woodward, G., Zegura, B. & Zimmermann, J. (2016) DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. *Research Ideas and Outcomes*, **2**, e11321.

Legendre, P. & Birks, H.J.B. (2012) From Classical to Canonical Ordination. In: *Tracking Environmental Change Using Lake Sediments: Data Handling and Numerical Techniques*. Springer Netherlands, Dordrecht, pp. 201–248.

Legendre, P. & Legendre, L.F.J. (2012) *Numerical Ecology*. Elsevier.

Legendre, P. (2014) Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, **23**, 1324–1334.

Lele, S.R., Keim, J.L. & Solymos, P. (2016) ResourceSelection: Resource Selection (Probability) Functions for Use-Availability Data. R package version 0.3-2.

Lemmens, P., Mergeay, J., De Bie, T., Van Wichelen, J., De Meester, L. & Declerck, S.A.J. (2013) How to maximally support local and regional biodiversity in applied conservation? Insights from pond management. *PLoS ONE*, **8**, e72538.

Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T. & 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**, 34.

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

Li, Y., Evans, N.T., Renshaw, M.A., Jerde, C.L., Olds, B.P., Shogren, A.J., Deiner, K., Lodge, D.M., Lamberti, G.A. & Pfrender, M.E. (2018a) Estimating fish alpha- and beta-diversity along a small stream with environmental DNA metabarcoding. *Metabarcoding and Metagenomics*, **2**, e24262.

Li, J., Lawson Handley, L.-J., Read, D.S. & Hänfling, B. (2018b) The effect of filtration method on the efficiency of environmental DNA capture and quantification via metabarcoding. *Molecular Ecology Resources*, **18**, 1102–1114.

Li, J., Lawson Handley, L., Harper, L.R., Brys, R., Watson, H.V. & Hänfling, B. (2018c) Limited dispersion and quick degradation of environmental DNA in fish ponds inferred by metabarcoding. *bioRxiv*, 459321.

Li, F., Peng, Y., Fang, W., Altermatt, F., Xie, Y., Yang, J. & Zhang, X. (2018d) Application of environmental DNA metabarcoding for predicting anthropogenic pollution in rivers. *Environmental Science & Technology*, **52**, 11708–11719.

Li, J., Hatton-Ellis, T.W., Lawson Handley, L.-J., Kimbell, H.S., Benucci, M., Peirson, G. & Hänfling, B. (2019) Ground-truthing of a fish-based environmental DNA metabarcoding method for assessing the quality of lakes. *Journal of Applied Ecology*.

Lobo, J., Shokralla, S., Costa, M.H., Hajibabaei, M. & Costa, F.O. (2017) DNA metabarcoding for high-throughput monitoring of estuarine macrobenthic communities. *Scientific Reports*, **7**, 15618.

Longcore, J.E., Pessier, A.P. & Nichols, D.K. (1999) Batrachochytrium Dendrobatidis gen. et sp. nov., a Chytrid Pathogenic to Amphibians. *Mycologia*, **91**, 219–227.

Lopes, C.M., Sasso, T., Valentini, A., Dejean, T., Martins, M., Zamudio, K.R. & Haddad, C.F.B. (2016) eDNA metabarcoding: a promising method for anuran surveys in highly diverse tropical forests. *Molecular Ecology Resources*, **17**, 904–914.

Lugg, W.H., Griffiths, J., van Rooyen, A.R., Weeks, A.R. & Tingley, R. (2017) Optimal survey designs for environmental DNA sampling. *Methods in Ecology and Evolution*, **9**, 1049–1059.

Lynx UK Trust (2018) Accessed on 1st December 2018. http://www.lynxuk.org/

Macan, T.T. (1960) *A key to the British fresh-and brackish-water gastropods: with notes on their ecology*. Freshwater Biological Association, Ambleside.

Maceda-Veiga, A., López, R. & Green, A.J. (2017) Dramatic impact of alien carp *Cyprinus carpio* on globally threatened diving ducks and other waterbirds in Mediterranean shallow lakes. *Biological Conservation*, **212**, 74–85.

Macher, J.-N. & Leese, F. (2017) Environmental DNA metabarcoding of rivers: Not all eDNA is everywhere, and not all the time. *bioRxiv*, 164046.

Macher, J.-N., Vivancos, A., Piggott, J.J., Centeno, F.C., Matthaei, C.D. & Leese, F. (2018) Comparison of environmental DNA and bulk-sample metabarcoding using highly degenerate cytochrome c oxidase I primers. *Molecular Ecology Resources*, **18**, 1456–1468.

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.

Magoč, T. & Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.

Maitland, P.S. (1972) *A key to the freshwater fishes of the British Isles with notes on their distribution and ecology*. Scientific Publication no. 27, Freshwater Biological Association, Ambleside, UK.

Maletzky, A., Kyek, M. & Goldschmid, A. (2007) Monitoring status, habitat features and amphibian species richness of crested newt (*Triturus cristatus* superspecies) ponds at the edge of the species range (Salzburg, Austria). *Annales de Limnologie - International Journal of Limnology*, **43**, 107–115.

Manchester, S.J. & Bullock, J.M. (2000) The impacts of non-native species on UK biodiversity and the effectiveness of control. *Journal of Applied Ecology*, **37**, 845–864.

Manenti, R. & Pennati, R. (2016) Environmental factors associated with amphibian breeding in streams and springs: effects of habitat and fish occurrence. *Amphibia-Reptilia*, **37**, 237–242.

Mantyka-Pringle, C.S., Martin, T.G., Moffatt, D.B., Linke, S. & Rhodes, J.R. (2014) Understanding and predicting the combined effects of climate change and land-use change on freshwater macroinvertebrates and fish. *Journal of Applied Ecology*, **51**, 572–581.

Marklund, O., Sandsten, H., Hansson, L.-A. & Blindow, I. (2002) Effects of waterfowl and fish on submerged vegetation and macroinvertebrates. *Freshwater Biology*, **47**, 2049–2059.

Martel, A., Spitzen-van der Sluijs, A., Blooi, M., Bert, W., Ducatelle, R., Fisher, M.C., Woeltjes, A., Bosman, W., Chiers, K., Bossuyt, F. & Pasmans, F. (2013) *Batrachochytrium salamandrivorans* sp. nov. causes lethal chytridiomycosis in amphibians. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 15325–15329.

Massimino, D., Harris, S.J. & Gillings, S. (2018) Evaluating spatiotemporal trends in terrestrial mammal abundance using data collected during bird surveys. *Biological Conservation*, **226**, 153–167.

Mathews, F., Kubasiewicz, L.M., Gurnell, J., Harrower, C.A., McDonald, R.A. & Shore, R.F. (2018) *A Review of the Population and Conservation Status of British Mammals: Technical Summary*. Natural England, Peterborough.

Matsuhashi, S., Doi, H., Fujiwara, A., Watanabe, S. & Minamoto, T. (2016) Evaluation of the Environmental DNA Method for Estimating Distribution and Biomass of Submerged Aquatic Plants. *PLoS ONE*, **11**, e0156217.

Mauvisseau, Q., Coignet, A., Delaunay, C., Pinet, F., Bouchon, D. & Souty-Grosset, C. (2018) Environmental DNA as an efficient tool for detecting invasive crayfishes in freshwater ponds. *Hydrobiologia*, **805**, 163–175.

Mazerolle, M.J. (2016) AICcmodavg: Model Selection and Multimodel Inference Based on (Q)AIC(c). R package version 2.1-1.

McDonald, C.P., Rover, J.A., Stets, E.G. & Striegl, R.G. (2012) The regional abundance and size distribution of lakes and reservoirs in the United States and implications for estimates of global lake extent. *Limnology and Oceanography*, **57**, 597–606.

McLee, A.G. & Scaife, R.W. (1992) The colonisation by great crested newts (*Triturus cristatus*) of a water body following treatment with a piscicide to remove a large population of sticklebacks (*Gasterosteus aculeatus*). *Herpetological Bulletin*, **42**, 6–9.

McShea, W.J., Forrester, T., Costello, R., He, Z. & Kays, R. (2016) Volunteer-run cameras as distributed sensors for macrosystem mammal research. *Landscape Ecology*, **31**, 55–66.

Menetrey, N., Sager, L., Oertli, B. & Lachavanne, J.-B. (2005) Looking for metrics to assess the trophic state of ponds. Macroinvertebrates and amphibians. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **15**, 653–664.

Messager, M.L. & Olden, J.D. (2018) Individual-based models forecast the spread and inform the management of an emerging riverine invader. *Diversity & Distributions*, **24**, 1816–1829.

Meusnier, I., Singer, G.A.C., Landry, J.-F., Hickey, D.A., Hebert, P.D.N. & Hajibabaei, M. (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, **9**, 214.

Minamoto, T., Fukuda, M., Katsuhara, K.R., Fujiwara, A., Hidaka, S., Yamamoto, S., Takahashi, K. & Masuda, R. (2017) Environmental DNA reflects spatial and temporal jellyfish distribution. *PLoS ONE*, **12**, e0173073.

Miralles, L., Dopico, E., Devlo-Delva, F. & Garcia-Vazquez, E. (2016) Controlling populations of invasive pygmy mussel (*Xenostrobus securis*) through citizen science and environmental DNA. *Marine Pollution Bulletin*, **110**, 127–132.

Mosher, B.A., Huyvaert, K.P. & Bailey, L.L. (2018) Beyond the swab: ecosystem sampling to understand the persistence of an amphibian pathogen. *Oecologia*, **188**, 319–330.

Muha, T.P., Rodríguez-Rey, M., Rolla, M. & Tricarico, E. (2017) Using Environmental DNA to Improve Species Distribution Models for Freshwater Invaders. *Frontiers in Ecology and Evolution*, **5**, 158.

Nakagawa, S. & Cuthill, I.C. (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews of the Cambridge Philosophical Society*, **82**, 591–605.

Nakagawa, H., Yamamoto, S., Sato, Y., Sado, T., Minamoto, T. & Miya, M. (2018) Comparing local- and regional-scale estimations of the diversity of stream fish using eDNA metabarcoding and conventional observation methods. *Freshwater Biology*, **63**, 569–580.

Nathan, L.M., Simmons, M., Wegleitner, B.J., Jerde, C.L. & Mahon, A.R. (2014) Quantifying environmental DNA signals for aquatic invasive species across multiple detection platforms. *Environmental Science & Technology*, **48**, 12800–12806.

National Biodiversity Network Atlas (2018) Accessed on 1st December 2018. https://nbnatlas.org/

Natural England (2015) Great crested newts: surveys and mitigation for development projects. Accessed on 21st November 2018. https://www.gov.uk/guidance/great-crested-newts-surveys-and-mitigation-for-development-projects

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.

Niemiller, M.L., Porter, M.L., Keany, J., Gilbert, H., Fong, D.W., Culver, D.C., Hobson, C.S., Denise Kendall, K., Davis, M.A. & Taylor, S.J. (2017) Evaluation of eDNA for groundwater invertebrate detection and monitoring: a case study with endangered *Stygobromus* (Amphipoda: Crangonyctidae). *Conservation Genetics Resources*, **10**, 247–257.

O'Donnell, J.L., Kelly, R.P., Lowell, N.C. & Port, J.A. (2016) Indexed PCR Primers Induce Template-Specific Bias in Large-Scale DNA Sequencing Studies. *PLoS ONE*, **11**, e0148698.

Oertli, B., Joye, D.A., Castella, E., Juge, R., Cambin, D. & Lachavanne, J.-B. (2002) Does size matter? The relationship between pond area and biodiversity. *Biological Conservation*, **104**, 59–70.

Oertli, B., Céréghino, R., Hull, A. & Miracle, R. (2009) Pond conservation: from science to practice. *Hydrobiologia*, **634**, 1–9.

Oldham, R.S., Keeble, J., Swan, M.J.S. & Jeffcote, M. (2000) Evaluating the suitability of habitat for the great crested newt (*Triturus cristatus*). *Herpetological Journal*, **10**, 143–155.

Olds, B.P., Jerde, C.L., Renshaw, M.A., Li, Y., Evans, N.T., Turner, C.R., Deiner, K., Mahon, A.R., Brueseke, M.A., Shirey, P.D., Pfrender, M.E., Lodge, D.M. & Lamberti, G.A. (2016) Estimating species richness using environmental DNA. *Ecology and Evolution*, **6**, 4214–4226.

Padgett-Stewart, T.M., Wilcox, T.M., Carim, K.J., McKelvey, K.S., Young, M.K. & Schwartz, M.K. (2016) An eDNA assay for river otter detection: a tool for surveying a semi-aquatic mammal. *Conservation Genetics Resources*, **8**, 5–7.

Paillisson, J.-M. & Marion, L. (2001) Interaction between coot (*Fulica atra*) and waterlily (*Nymphaea alba*) in a lake: the indirect impact of foraging. *Aquatic Botany*, **71**, 209–216.

Paradis, E. & Schliep, K. (2018) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 1–3.

Parsons, K.M., Everett, M., Dahlheim, M. & Park, L. (2018) Water, water everywhere: environmental DNA can unlock population structure in elusive marine species. *Royal Society Open Science*, **5**, 180537.

Peres-Neto, P.R., Legendre, P., Dray, S. & Borcard, D. (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, **87**, 2614–2625.

Perrow, M.R., Schutten, H.J., Howes, J.R., Holzer, T., Madgwick, F.J. & Jowitt, A. (1997) Interactions between coot (*Fulica atra*) and submerged macrophytes: the role of birds in the restoration process. *Hydrobiologia*, **342**, 241–255.

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.

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.

Piha, H., Luoto, M. & Merilä, J. (2007) Amphibian occurrence is influenced by current and historic landscape characteristics. *Ecological Applications*, **17**, 2298–2309.

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.

Pittock, J., Hansen, L.J. & Abell, R. (2008) Running dry: Freshwater biodiversity, protected areas and climate change. *Biodiversity*, **9**, 30–38.

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.

Preston, D.L., Hedman, H.D., Esfahani, E.R., Pena, E.M., Boland, C.E., Lunde, K.B. & Johnson, P.T.J. (2017) Responses of a wetland ecosystem to the controlled introduction of invasive fish. *Freshwater Biology*, **62**, 767–778.

Qu, C. & Stewart, K.A. (2019) Evaluating monitoring options for conservation: comparing traditional and environmental DNA tools for a critically endangered mammal. *The Science of Nature*, **106**, 9.

R Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria. Available from: https://www.R-project.org

Raemy, M. & Ursenbacher, S. (2018) Detection of the European pond turtle (*Emys orbicularis*) by environmental DNA: is eDNA adequate for reptiles? *Amphibia-Reptilia*, **39**, 135–143.

Rannap, R. & Briggs, L. (2006). THE CHARACTERISTICS OF GREAT CRESTED NEWT TRITURUS CRISTATUS' BREEDING PONDS. Protection of the Great Crested Newt. Project Report. The experiences of LIFE-Nature project "Protection of *Triturus cristatus* in the Eastern Baltic Region" LIFE04NAT/EE/000070. Ministry of the Environment of the Republic of Estonia, Tallinn.

Rannap, R., Lõhmus, A. & Briggs, L. (2009a) Niche position, but not niche breadth, differs in two coexisting amphibians having contrasting trends in Europe. *Diversity and Distributions*, **15**, 692–700.

Rannap, R., Lõhmus, A. & Briggs, L. (2009b) Restoring ponds for amphibians: a success story. *Hydrobiologia*, **634**, 87–95.

Reading, C.J. & Jofré, G.M. (2009) Habitat selection and range size of grass snakes *Natrix natrix* in an agricultural landscape in southern England. *Amphibia-Reptilia*, **30**, 379–388.

Rees, H.C., Bishop, K., Middleditch, D.J., Patmore, J.R.M., Maddison, B.C. & Gough, K.C. (2014a) The application of eDNA for monitoring of the Great Crested Newt in the UK. *Ecology and Evolution*, **4**, 4023–4032.

Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R.M. & Gough, K.C. (2014b) REVIEW: The detection of aquatic animal species using environmental DNA–a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, **51**, 1450–1459.

Rees, H.C., Baker, C.A., Gardner, D.S., Maddison, B.C. & Gough, K.C. (2017) The detection of great crested newts year round via environmental DNA analysis. *BMC Research Notes*, **10**, 1–4.

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

Rice, C.N. (2016). *Abundance, impacts and resident perceptions of non-native common pheasants (Phasianus colchicus) in Jersey, UK Channel Islands*. Masters Thesis, University of Kent.

Robinson, C.V., Uren Webster, T.M., Cable, J., James, J. & Consuegra, S. (2018) Simultaneous detection of invasive signal crayfish, endangered white-clawed crayfish and the crayfish plague pathogen using environmental DNA. *Biological Conservation*, **222**, 241–252.

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.

Rodgers, T.W. & Mock, K.E. (2015) Drinking water as a source of environmental DNA for the detection of terrestrial wildlife species. *Conservation Genetics Resources*, **7**, 693–696.

Rodgers, T. (2017) Proper fin-clip sample collection for molecular analyses in the age of eDNA. *Journal of Fish Biology*, **91**, 1265–1267.

Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, **4**, e2584.

Rovero, F., Martin, E., Rosa, M., Ahumada, J.A. & Spitale, D. (2014) Estimating species richness and modelling habitat preferences of tropical forest mammals from camera trap data. *PLoS ONE*, **9**, e103300.

Roy, M., Belliveau, V., Mandrak, N.E. & Gagné, N. (2018) Development of environmental DNA (eDNA) methods for detecting high-risk freshwater fishes in live trade in Canada. *Biological Invasions*, **20**, 299-314.

Ruane, N.M., Davenport, J. & Igoe, F. (2012) NON-DESTRUCTIVE TECHNIQUES FOR THE MONITORING OF ARCTIC CHAR *SALVELINUS ALPINUS* (L.) IN IRISH LOUGHS I. FYKE NETTING. *Biology and Environment: Proceedings of the Royal Irish Academy*, **112B**, 301–304.

Ruiz-Olmo, J. & Gosálbez, J. (1997) Observations on the sprainting behaviour of the otter *Lutra lutra* in the NE Spain. *Acta Theriologica*, **42**, 259–270.

Sadlier, L.M.J., Webbon, C.C., Baker, P.J. & Harris, S. (2004) Methods of monitoring red foxes *Vulpes vulpes* and badgers *Meles meles*: are field signs the answer? *Mammal Review*, **34**, 75–98.

Sala, O.E., Chapin, F.S., III, Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., Leemans, R., Lodge, D.M., Mooney, H.A., Oesterheld, M., Poff, N.L., Sykes, M.T., Walker, B.H., Walker, M. & Wall, D.H. (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770–1774.

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*, **13**, e0192409.

Sasso, T., Lopes, C.M., Valentini, A., Dejean, T., Zamudio, K.R., Haddad, C.F.B. & Martins, M. (2017) Environmental DNA characterization of amphibian communities in the Brazilian Atlantic forest: Potential application for conservation of a rich and threatened fauna. *Biological conservation*, **215**, 225–232.

Savage, A.A. (1989) *Adults of the British aquatic Hemiptera Heteroptera: a key with ecological notes*. Freshwater Biological Association, Ambleside.

Sayer, C.D., Copp, G.H., Emson, D., Godard, M.J., Zięba, G. & Wesley, K.J. (2011) Towards the conservation of crucian carp *Carassius carassius*: understanding the extent and causes of decline within part of its native English range. *Journal of Fish Biology*, **79**, 1608–1624.

Sayer, C., Andrews, K., Shilland, E., Edmonds, N., Edmonds-Brown, R., Patmore, I., Emson, D. & Axmacher, J. (2012) The role of pond management for biodiversity conservation in an agricultural landscape. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **22**, 626–638.

Sayer, C., Shilland, E., Greaves, H., Dawson, B., Patmore, I., Emson, D., Alderton, E., Robinson, P., Andrews, K., Axmacher, J. & Wiik, E. (2013) Managing Britain's ponds - conservation lessons from a Norfolk farm. *British Wildlife*, **25**, 21–28.

Schilling, E.G., Loftin, C.S. & Huryn, A.D. (2009a) Effects of introduced fish on macroinvertebrate communities in historically fishless headwater and kettle lakes. *Biological Conservation*, **142**, 3030–3038.

Schilling, E.G., Loftin, C.S. & Huryn, A.D. (2009b) Macroinvertebrates as indicators of fish absence in naturally fishless lakes. *Freshwater Biology*, **54**, 181–202.

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

Schmidt, B.R., Kéry, M., Ursenbacher, S., Hyman, O.J. & 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**, 646–653.

Schneider, J., Valentini, A., Dejean, T., Montarsi, F., Taberlet, P., Glaizot, O. & Fumagalli, L. (2016) Detection of Invasive Mosquito Vectors Using Environmental DNA (eDNA) from Water Samples. *PLoS ONE*, **11**, e0162493.

Schnell, I.B., Thomsen, P.F., Wilkinson, N., Rasmussen, M., Jensen, L.R.D., Willerslev, E., Bertelsen, M.F. & Gilbert, M.T.P. (2012) Screening mammal biodiversity using DNA from leeches. *Current Biology*, **22**, R262–3.

Schnell, I.B., Bohmann, K. & Gilbert, M.T.P. (2015) Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, **15**, 1289–1303.

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

Scriver, M., Marinich, A., Wilson, C. & Freeland, J. (2015) Development of species-specific environmental DNA (eDNA) markers for invasive aquatic plants. *Aquatic Botany*, **122**, 27–31.

Sellers, G.S., Di Muri, C., Gómez, A. & Hänfling, B. (2018) Mu-DNA: a modular universal DNA extraction method adaptable for a wide range of sample types. *Metabarcoding and Metagenomics*, **2**, e24556.

Sewell, D. & Griffiths, R. (2009) Can a Single Amphibian Species Be a Good Biodiversity Indicator? *Diversity*, **1**, 102–117.

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., Edwards, F., Emmett, B.A., Bik, H.M. & Creer, S. (2018) Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms. *Communications Biology*, **1**, 1–8.

Shaw, J.L.A., Clarke, L.J., Wedderburn, S.D., Barnes, T.C., Weyrich, L.S. & Cooper, A. (2016a) Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biological Conservation*, **197**, 131–138.

Shaw, J.L.A., Weyrich, L. & Cooper, A. (2016b) Using environmental (e)DNA sequencing for aquatic biodiversity surveys: a beginner's guide. *Marine and Freshwater Research*, **68**, 20–33.

Sheehy, E., Sutherland, C., O'Reilly, C. & Lambin, X. (2018) The enemy of my enemy is my friend: native pine marten recovery reverses the decline of the red squirrel by suppressing grey squirrel populations. *Proceedings of the Royal Society of London B*, **285**, 20172603.

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.

Sigsgaard, E.E., Nielsen, I.B., Bach, S.S., Lorenzen, E.D., Robinson, D.P., Knudsen, S.W., Pedersen, M.W., Al Jaidah, M., Orlando, L., Willerslev, E., Møller, P.R. & Thomsen, P.F. (2016) Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution*, **1**, 0004.

Simpfendorfer, C.A., Kyne, P.M., Noble, T.H., Goldsbury, J., Basiita, R.K., Lindsay, R., Shields, A., Perry, C. & Jerry, D.R. (2016) Environmental DNA detects Critically Endangered largetooth sawfish in the wild. *Endangered Species Research*, **30**, 109–116.

Skei, J.K., Dolmen, D., Rønning, L. & Ringsby, T.H. (2006) Habitat use during the aquatic phase of the newts *Triturus vulgaris* (L.) and *T. cristatus* (Laurenti) in central Norway: proposition for a conservation and monitoring area. *Amphibia-Reptilia*, **27**, 309–324.

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.

Smart, A.S., Weeks, A.R., van Rooyen, A.R., Moore, A., McCarthy, M.A. & Tingley, R. (2016) Assessing the cost-efficiency of environmental DNA sampling. *Methods in Ecology and Evolution*, **7**, 1291–1298.

Smith, P. & Moss, B. (1994) *The role of fish in the management of freshwater Sites of Special Scientific Interest*. English Nature Research Report no. 111.

de Souza, L.S., Godwin, J.C., Renshaw, M.A. & Larson, E. (2016) Environmental DNA (eDNA) Detection Probability Is Influenced by Seasonal Activity of Organisms. *PLoS ONE*, **11**, e0165273.

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.

Spens, J., Evans, A.R., Halfmaerten, D., Knudsen, S.W., Sengupta, M.E., Mak, S.S.T., Sigsgaard, E.E. & Hellström, M. (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**, 635–645.

Stat, M., John, J., DiBattista, J.D., Newman, S.J., Bunce, M. & Harvey, E.S. (2019) Combined use of eDNA metabarcoding and video surveillance for the assessment of fish biodiversity. *Conservation Biology*, **33**, 196–205.

Stefanoudis, P.V., Sayer, C.D., Greaves, H.M., Davidson, T.A., Robson, H., Almeida, D. & Smith, E. (2017) Consequences of fish for cladoceran, water beetle and macrophyte communities in a farmland pond landscape: implications for conservation. *Fundamental and Applied Limnology*, **190**, 141–156.

Stoeckle, B.C., Beggel, S., Cerwenka, A.F., Motivans, E., Kuehn, R. & Geist, J. (2017) A systematic approach to evaluate the influence of environmental conditions on eDNA detection success in aquatic ecosystems. *PLoS ONE*, **12**, e0189119.

Strand, D.A., Jussila, J., Johnsen, S.I., Viljamaa-Dirks, S., Edsman, L., Wiik-Nielsen, J., Viljugrein, H., Engdahl, F. & Vrålstad, T. (2014) Detection of crayfish plague spores in large freshwater systems. *Journal of Applied Ecology*, **51**, 544–553.

Strayer, D.L. & Dudgeon, D. (2010) Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society*, **29**, 344–358.

Strickland, G.J. & Roberts, J.H. (2019) Utility of eDNA and occupancy models for monitoring an endangered fish across diverse riverine habitats. *Hydrobiologia*, **826**, 129–144.

Strickler, K.M., Fremier, A.K. & Goldberg, C.S. (2015) Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation*, **183**, 85–92.

Stuart, S.N., Chanson, J.S., Cox, N.A., Young, B.E., Rodrigues, A.S.L., Fischman, D.L. & Waller, R.W. (2004) Status and trends of amphibian declines and extinctions worldwide. *Science*, **306**, 1783–1786.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, **21**, 2045–2050.

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

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

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.

Takahashi, M.K., Meyer, M.J., Mcphee, C., Gaston, J.R., Venesky, M.D. & Case, B.F. (2018) Seasonal and diel signature of eastern hellbender environmental DNA. *The Journal of Wildlife Management*, **82**, 217–225.

Tarkan, A.S., Copp, G.H., Zięba, G., Godard, M.J. & Cucherousset, J. (2009) Growth and reproduction of threatened native crucian carp *Carassius carassius* in small ponds of Epping Forest, south-east England. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **19**, 797–805.

Tessler, M., Weiskopf, S.R., Berniker, L., Hersch, R., Mccarthy, K.P., Yu, D.W. & Siddall, M.E. (2018) Bloodlines: mammals, leeches, and conservation in southern Asia. *Systematics and Biodiversity*, **16**, 488–496.

Thomsen, P.F., 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.

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

Thomsen, P.F., Møller, P.R., Sigsgaard, E.E., Knudsen, S.W., Jørgensen, O.A. & Willerslev, E. (2016) Environmental DNA from Seawater Samples Correlate with Trawl Catches of Subarctic, Deepwater Fishes. *PLoS ONE*, **11**, 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**, 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.

Trevelline, B.K., Nuttle, T., Hoenig, B.D., Brouwer, N.L., Porter, B.A. & Latta, S.C. (2018) DNA metabarcoding of nestling feces reveals provisioning of aquatic prey and resource partitioning among Neotropical migratory songbirds in a riparian habitat. *Oecologia*, **187**, 85–98.

Tsuji, S., Ushio, M., Sakurai, S., Minamoto, T. & Yamanaka, H. (2017) Water temperature-dependent degradation of environmental DNA and its relation to bacterial abundance. *PLoS ONE*, **12**, e0176608.

Turner, C.R., Lodge, D.M., Xu, C., Cooper, M.J. & Lamberti, G.A. (2012) *Evaluating environmental DNA detection alongside standard fish sampling in Great Lakes coastal wetland monitoring (seed project)*. Final Report, Illinois-Indiana Sea Grant.

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., 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., 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.

Unglaub, B., Steinfartz, S., Drechsler, A. & Schmidt, B.R. (2015) Linking habitat suitability to demography in a pond-breeding amphibian. *Frontiers in Zoology*, **12**, 9.

Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., Murata, K., Nikaido, M., Sado, T., Sato, Y., Takeshita, M., Iwasaki, W., Yamanaka, H., Kondoh, M. & Miya, M. (2017) Environmental DNA enables detection of terrestrial mammals from forest pond water. *Molecular Ecology Resources*, **17**, e63–e75.

Ushio, M., Murakami, H., Masuda, R., Sado, T., Miya, M., Sakurai, S., Yamanaka, H., Minamoto, T. & Kondoh, M. (2018a) Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. *Metabarcoding and Metagenomics*, **2**, e23297.

Ushio, M., Murata, K., Sado, T., Nishiumi, I., Takeshita, M., Iwasaki, W. & Miya, M. (2018b) Demonstration of the potential of environmental DNA as a tool for the detection of avian species. *Scientific Reports*, **8**, 4493.

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

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G.H., Geniez, P., Pont, D., Argillier, C., Baudoin, J.-M., Peroux, T., Crivelli, A.J., Olivier, A., Acqueberge, M., Le Brun, M., Møller, P.R., Willerslev, E. & Dejean, T. (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, **25**, 929–942.

Vamos, E.E., Elbrecht, V. & Leese, F. (2017) Short COI markers for freshwater macroinvertebrate metabarcoding. *Metabarcoding and Metagenomics*, **1**, e14625.

Vences, M., Lyra, M.L., Perl, R.G.B., Bletz, M.C., Stanković, D., Lopes, C.M., Jarek, M., Bhuju, S., Geffers, R., Haddad, C.F.B. & Steinfartz, S. (2016) Freshwater vertebrate metabarcoding on Illumina platforms using double-indexed primers of the mitochondrial 16S rRNA gene. *Conservation Genetics Resources*, **8**, 323–327.

Vuorio, V., Heikkinen, R.K. & Tikkanen, O.-P. (2013) Breeding Success of the Threatened Great Crested Newt in Boreal Forest Ponds. *Annales Zoologici Fennici*, **50**, 158–169.

Walker, S.F., Salas, M.B., Jenkins, D., Garner, T.W.J., Cunningham, A.A., Hyatt, A.D., Bosch, J. & Fisher, M.C. (2007) Environmental detection of *Batrachochytrium dendrobatidis* in a temperate climate. *Diseases of Aquatic Organisms*, **77**, 105–112.

Wallace I.D., Wallace B. & Philipson G.N. (1990) *A key to the case-bearing caddis larvae of Britain and Ireland*. Freshwater Biological Association, Ambleside.

Wallau, G.L., Della-Flora, F., Bueno, A.S., Corso, J., Ortiz, M.F. & Cáceres, N.C. (2010) Behaviour of the Common Moorhen in Rio Grande do Sul, Brazil. *Acta Ethologica*, **13**, 127–135.

Weltz, K., Lyle, J.M., Ovenden, J., Morgan, J.A.T., Moreno, D.A. & Semmens, J.M. (2017) Application of environmental DNA to detect an endangered marine skate species in the wild. *PLoS ONE*, **12**, e0178124.

Wheat, R.E., Allen, J.M., Miller, S.D.L., Wilmers, C.C. & Levi, T. (2016) Environmental DNA from Residual Saliva for Efficient Noninvasive Genetic Monitoring of Brown Bears (*Ursus arctos*). *PLoS ONE*, **11**, e0165259.

Wheeler, A. (1977) The Origin and Distribution of the Freshwater Fishes of the British Isles. *Journal of Biogeography*, **4**, 1–24.

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer, New York, USA. http://ggplot2.org

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

Wilcox, T.M., McKelvey, K.S., Young, M.K., Sepulveda, A.J., Shepard, B.B., Jane, S.F., Whiteley, A.R., Lowe, W.H. & Schwartz, M.K. (2016) Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char *Salvelinus fontinalis*. *Biological Conservation*, **194**, 209–216.

Wilkinson, S.P., Davy, S.K., Bunce, M. & Stat, M. (2018) Taxonomic Identification of Environmental DNA with Informatic Sequence Classification Trees. *PeerJ Preprints*.

Williams, K.E., Huyvaert, K.P. & Piaggio, A.J. (2017) Clearing muddied waters: Capture of environmental DNA from turbid waters. *PLoS ONE*, **12**, e0179282.

Williams, K.E., Huyvaert, K.P., Vercauteren, K.C., Davis, A.J. & Piaggio, A.J. (2018) Detection and persistence of environmental DNA from an invasive, terrestrial mammal. *Ecology and Evolution*, **8**, 688–695.

Williams, P., Whitfield, M., Biggs, J., Bray, S., Fox, G., Nicolet, P. & Sear, D. (2003) Comparative biodiversity of rivers, streams, ditches and ponds in an agricultural landscape in Southern England. *Biological Conservation*, **115**, 329–341.

Williams, P., Whitfield, M. & Biggs, J. (2008) How can we make new ponds biodiverse? A case study monitored over 7 years. *Hydrobiologia*, **597**, 137–148.

Williams, P., Biggs, J., Crowe, A., Murphy, J., Nicolet, P., Weatherby, A. & Dunbar, M. (2010) Countryside Survey: Ponds Report from 2007. Accessed on 1st December 2018. https://countrysidesurvey.org.uk/content/ponds-report-2007

Wood, P.J., Greenwood, M.T., Barker, S.A. & Gunn, J. (2001) The effects of amenity management for angling on the conservation value of aquatic invertebrate communities in old industrial ponds. *Biological Conservation*, **102**, 17–29.

Wood, P.J., Greenwood, M.T. & Agnew, M.D. (2003) Pond biodiversity and habitat loss in the UK. *Area*, **35**, 206–216.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T.L. (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, **13**, 1-11.

Zeale, M.R.K., Butlin, R.K., Barker, G.L.A., Lees, D.C. & Jones, G. (2011) Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*, **11**, 236–244.

Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, **7**, 203–214.

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

Zizka, V.M.A., Leese, F., Peinert, B. & Geiger, M.F. (2018) DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method. *Genome*.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A. & Smith, G.M. (2009) *Mixed effects models and extensions in ecology with R*. Springer, New York, USA.

Appendices

Common coot (*Fulica atra*) (Linnaeus, 1758) feeding chick a water beetle

© user: Tore Bustad | Flickr | CC BY-NC-ND 2.0

Appendix 2

Appendix 2.1: Supplementary methods

Samples

In accordance with the eDNA sampling methodology outlined by Biggs *et al*. (2015), 20 x 30 mL water samples were collected at even intervals around the pond margin and pooled in a sterile 1 L Whirl-Pak® stand-up bag, which was shaken to provide a single homogenised sample from each pond. Six 15 mL subsamples were taken from the mixed sample using a sterile plastic pipette (25 mL) and added to sample tubes, containing 33.5 mL absolute ethanol and 1.5 mL sodium acetate 3 M (pH 5.2), for ethanol precipitation. Subsamples were then sent to Fera Science Ltd (Natural England) and ADAS (private contracts) for eDNA analysis according to laboratory protocols established by Biggs *et al*. (2015). Subsamples were centrifuged at 14,000 x g for 30 minutes at 6 ^{o}C and the supernatant discarded. Subsamples were then pooled during the first step of DNA extraction with the DNeasy Blood & Tissue Kit® (Qiagen®, Hilden, Germany), where 360 µL of ATL buffer was added to the first tube, vortexed, and the supernatant transferred to the second tube. This process was repeated for all six tubes. The supernatant in the sixth tube, containing concentrated DNA from all six subsamples, was transferred in a 2 mL tube and extraction continued following manufacturer's instructions to produce one eDNA sample per pond.

Targeted quantitative PCR (qPCR)

Prior to testing for great crested newt (*Triturus cristatus*), all extracted samples were tested for PCR inhibitors and sample degradation using the methodology outlined by Biggs *et al.* (2015), where an Internal Positive Control was included in qPCR reactions of eDNA samples and a sample considered inhibited if replicates showed different C_q values (where samples move into the exponential phase of qPCR amplification). Targeted qPCR was carried out as part of the *T. cristatus* monitoring programmes mentioned above in the laboratories at Fera Science Ltd and ADAS during 2015 using a standardised protocol (Biggs *et al.*, 2015). Extracted DNA was amplified by TaqMan probe qPCR using published primers and probe to amplify an 81 bp fragment of the cytochrome *b*

gene: TCCBL (5’-CGTAAACTACGGCTGACTAGTACGAA-3’), TCCBR (5-CCGATGTGTATGTAGATGCAAACA) and TCCB_Probe (5’-CCACGCTAACGGAGCCTCGC-3’) (Thomsen *et al.*, 2012). PCR reactions were set up in a total volume of 25 µL consisting of: 3 µL of extracted template DNA, 1 µL of each primer (0.4 µM), 1 µL of probe (0.1 µM), 1x TaqMan® Environmental Master Mix 2.0 (containing AmpliTaq GOLD DNA polymerase, Life Technologies) and ddH_2O. The PCR included an initial incubation for 5 min at 50 °C, then a 10 min denaturation step at 95 °C, followed by 55 cycles of denaturation at 95 °C for 30 s and annealing at 56.3 °C for 1 min. For each sample, 12 qPCR replicates were performed and a sample recorded as positive for *T. cristatus* if one or more qPCR replicates were positive. Positive (*T. cristatus* DNA: 1×10^{-1} ng/µL to 1×10^{-4} ng/µL) and negative controls (ddH_2O) were also included on each plate in quadruplicate. Following qPCR, the eDNA samples were placed in storage at -80 °C.

Reference database construction

A custom, phylogenetically curated reference database of the target region was created for UK vertebrate species. For freshwater fish, we used a previously created database comprising 67 fish species, which includes all known native and non-native species in the UK and our positive control *Rhamphochromis esox*, a species of cichlid from Lake Malawi (Hänfling *et al.* 2016). For all remaining vertebrate species recorded in the UK (Natural History Museum UK Species Database, 2017), custom, phylogenetically curated reference databases were constructed using the ReproPhylo environment (Szitenberg *et al.*, 2015) in a Jupyter notebook (Kluyver *et al.*, 2016). Database curation for each of the main UK vertebrate groups (amphibians, birds, mammals, reptiles) was performed separately to ease data processing. Jupyter notebooks detailing the processing steps for each data subset are deposited in a dedicated GitHub repository for Chapter 2 (https://github.com/HullUni-bioinformatics/Harper_et_al_2018), which has been permanently archived (https://doi.org/10.5281/zenodo.2633978). Species lists containing the binomial nomenclature of UK vertebrate species were constructed using the Natural History Museum UK Species Database. All vertebrates recorded in the UK were included. The BioPython script performed a GenBank search based on the species lists and downloaded all available mitochondrial 12S ribosomal RNA (rRNA) sequences for specified species. Proportion of reference sequences available for species varied within

each vertebrate group: amphibians 100.00% (*N* = 21), reptiles 90.00% (*N* = 20), birds 55.88% (*N* = 621), and mammals (83.93%, *N* = 112). Where there were no records on GenBank for a UK species, the database was supplemented with downloaded sequences belonging to sister species in the same genus. Species that had no 12S rRNA records on Genbank are provided in Table S2.1.

Redundant sequences were removed by clustering at 100% similarity using vsearch 1.1 (Rognes *et al.*, 2016). Due to high proportion of partial 12S rRNA records on GenBank for the majority of UK species, only sequences longer than 500 bp were processed initially to increase alignment robustness to large gaps. Sequences were aligned using MUSCLE (Edgar, 2004). Short sequences can cause problems in global paired alignments where the alignment algorithm attempts to align them to longer sequences. Short 12S rRNA sequences (<500 bp) were later incorporated into the existing long 12S rRNA alignment using the hmmer v3 program suite (HMMER development team, 2016) to construct a Hidden Markov Model alignment containing sequences of all lengths. Alignments were trimmed using trimAl (Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009). Maximum likelihood trees were inferred with RAxML 8.0.2 (Stamatakis, 2006) using the GTR+gamma model of substitutions. The complete alignments were then processed using SATIVA (Kozlov *et al.*, 2016) for automated identification of 'mislabelled' sequences which could cause conflict in downstream analyses. Putatively mislabelled sequences were removed and process of alignment and phylogenetic tree construction repeated for manual investigation of sequences. The resultant databases (i.e. curated, non-redundant reference databases) contained: 198 amphibian sequences from 20/21 species, 112 reptile sequences from 19/20 species, 272 fish sequences from 60/62 species, 940 mammal sequences from 95/112 species, and 622 bird sequences from 347/621 species. Databases for each vertebrate group were concatenated and the combined vertebrate database used for *in silico* validation of primers.

The amphibian database was supplemented by Sanger sequences obtained from tissue of *T. cristatus*, smooth newt (*Lissotriton vulgaris*), Alpine newt (*Ichthyosaura alpestris*), common toad (*Bufo bufo*), which were supplied by the University of Kent under licence from Natural England, and common frog (*Rana temporaria*), supplied by the University of Glasgow. Amphibian DNA from the University of Kent was extracted from tissue samples using a DNeasy Blood & Tissue Kit® (Qiagen®, Hilden, Germany) under licence from Natural England by H. Rees. Reference sequences of the entire 12S rRNA region were generated by three sets of novel primers:

***T. cristatus* (61 °C)**: Newt_F1 5'-GCACTGAAAATGCTAAGACAGA-3'
Newt_R6 5'-CAGGTATTTTCTCGGTGTAAGCA-3'

Newts (59 °C): Newt_F2 5'-GCACTGAAAATGCTAAGACAG-3'
Newt_R1 5'-TCTCGGTGTAAGCAAGATGC-3'

Anura (57 °C): AnuraShort_F2 5'-TCCACTGGTCTTAGGAGCCA-3'
AnuraShort_R1 5'-ACCATGTTACGACTTGCCTC-3'

Primers were designed from an alignment of tRNA, 12S and 16S rRNA regions in UK Caudata and Anura species. PCR reactions were performed in 25 µL volumes containing: 12.5 µL of MyTaq™ Red Mix (Bioline®, UK), 1 µL (final concentration - 0.04 µM) of forward and reverse primer (Integrated DNA Technologies, Belgium), 8.5µL of molecular grade sterile water (Fisher Scientific UK Ltd, UK) and 2 µL DNA template. PCRs were performed on an Applied Biosystems® Veriti Thermal Cycler (Fisher Scientific UK Ltd, UK) with the following profile: 95 °C for 3 min, 35 cycles of 95 °C for 30 sec, x °C (see temperatures above) for 60 sec and 72 °C for 90 sec, followed by a final elongation step at 72 °C for 10 min. Purified PCR products were Sanger sequenced directly (Macrogen Europe, Amsterdam, Netherlands) in both directions using the PCR primers. Sequences were edited using CodonCode Aligner (CodonCode Corporation, MA, USA). The complete reference database compiled in GenBank format has been deposited in the GitHub repository for Chapters 2 and 3.

Primer validation

Vertebrate DNA from eDNA samples was amplified with published 12S rRNA primers 12S-V5-F (5'-ACTGGGATTAGATACCCC-3') and 12S-V5-R (5'-TAGAACAGGCTCCTCTAG-3') (Riaz *et al*., 2011). Primers were validated for the present study *in silico* using ecoPCR software (Ficetola *et al*., 2010) against a custom, phylogenetically curated reference database for UK vertebrates. Parameters were set to allow a fragment size of 50-250 bp and maximum of three mismatches between the primer pair and each sequence in the reference database. Primers were previously validated *in vitro* for UK fish communities by Hänfling *et al*. (2016) and here were also validated against tissue DNA extracted from UK amphibian species: *T. cristatus*, *L. vulgaris*, palmate newt (*Lissotrition helveticus*), *I. alpestris*, *R. temporaria* and *B. bufo*. Primer

validation tests were performed at the University of Hull in a separate laboratory situated on a different floor to the dedicated eDNA laboratory. A dilution series (10^0 to 10^{-8}) was performed for DNA (standardised to 5 ng/µL) from each species to identify the limit of detection (LOD) for each species. Molecular grade sterile water (Fisher Scientific UK Ltd, UK) substituted template DNA for the PCR negative control.

eDNA metabarcoding

A two-step PCR protocol was performed on eDNA samples at the University of Hull. Dedicated rooms were available for pre-PCR and post-PCR processes. Pre-PCR processes were performed in a dedicated eDNA laboratory, with separate rooms for filtration, DNA extraction and PCR preparation of sensitive environmental samples. PCR reactions were set up in an ultraviolet and bleach sterilized laminar flow hood. Eight-strip PCR tubes with individually attached lids were used instead of 96-well plates to minimise cross-contamination risk between samples (Port *et al*., 2016). After the first sequencing run revealed substantial human (*Homo sapiens*) DNA contamination across samples and PCR controls, reactions prepared for the second sequencing run were sealed with mineral oil as an additional measure against PCR contamination. For the first PCR, three replicates were performed for each sample to combat PCR stochasticity. Alternating PCR positive and negative controls were included on each PCR strip (six positive and six negative controls on each 96-well plate), to screen for sources of potential contamination. The DNA used for the PCR positive control was *R. esox,* as occurrence in UK ponds is extremely rare or non-existent. The negative control substituted molecular grade sterile water (Fisher Scientific UK Ltd, UK) for template DNA.

During the first PCR, the target region was amplified using the primers described above, including adapters (Illumina, 2011). First step PCR reactions were performed in a final volume of 21.1 µL, using 2 µL of DNA extract as a template. The amplification mixture contained 10.5 µL of MyTaq™ HS Red Mix (Bioline®, UK), 1.05 µL (final concentration - 0.5 µM) of forward and reverse primer (Integrated DNA Technologies, Belgium) and 6.5 µL of molecular grade sterile water (Fisher Scientific UK Ltd, UK). PCR was performed on an Applied Biosystems® Veriti Thermal Cycler (Fisher Scientific UK Ltd, UK) and PCR conditions for the first component of the two-step protocol consisted of: an incubation step at 98 °C for 5 min, followed by 35 cycles of denaturation at 98 °C for 15 s, annealing at 56 °C for 20 s, and extension at 72 °C for 30 s with final

extension at 72 °C for 10 min. PCR products were stored at 4 °C until fragment size was verified by visualising 5 µL of selected PCR products on 2% agarose gels (100 mL 0.5x TBE buffer, 2 g agarose powder). Gels were then stained with ethidium bromide and imaged using Image Lab Software (Bio-Rad Laboratories Ltd, UK). A PCR product was deemed positive where there was an amplification band on the gel that was of the expected size (200-300 bp). PCR replicates for each sample were pooled in preparation for the addition of Illumina indexes in the second PCR, which resulted in 63.3 µL of PCR product for each sample. PCR positive and negative controls were not pooled to allow individual purification and sequencing of all 228 PCR controls. All PCR products (30 µL samples and 15 µL PCR controls) were then purified to remove excess primer using E.Z.N.A® Cycle Pure V-Spin Clean-Up Kits (Omega Bio-tek, GA, USA) following manufacturers protocol. Eluted DNA was stored at -20 °C until the second PCR could be performed.

In the second PCR, Multiplex Identification (MID) tags (unique 8-nucleotide sequences) and Illumina MiSeq adapter sequences were bound to the amplified product. These tags were included in the forward and reverse primers resulting in indexed primers for second PCR (O'Donnell *et al.*, 2016). For each second PCR plate, 96 unique tag combinations were created by combining eight unique forward tags with 12 unique reverse tags or vice versa (Kitson *et al.*, 2019). A total of 384 unique tag combinations were achieved, allowing samples to be distinguished during bioinformatics analysis. Second step PCR reactions were performed in eight-strip PCR tubes with individually attached lids in a final volume of 21.1 µL, using 2 µL of purified DNA from the first PCR product as a template. The amplification mixture contained 10.5 µL of MyTaq™ HS Red Mix (Bioline®, UK), 2.1 µL (final concentration - 0.5 µM) of tagged primer mix (Integrated DNA Technologies, Belgium) and 6.5 µL of molecular grade sterile water (Fisher Scientific UK Ltd, UK). PCR was performed on an Applied Biosystems® Veriti Thermal Cycler (Fisher Scientific UK Ltd, UK) with the following profile: denaturation at 95 °C for 3 min, followed by 12 cycles of annealing at 98 °C for 20 s and extension at 72 °C for 30 s with final extension at 72 °C for 5 min. PCR products were stored at 4 °C before they were all visualised on 2% agarose gels (100 mL 0.5x TBE buffer, 2 g agarose powder) using 5 µL PCR product. Gels were then stained with ethidium bromide and imaged using Image Lab Software (Bio-Rad Laboratories Ltd, UK). Again, PCR products were deemed positive where there was an amplification band on the gel that was of the expected size (200-300 bp). Amplification bands were found to be present in some of the negative controls thus all negative controls were included for sequencing.

All remaining library preparation was conducted at Fera Science Ltd. PCR products were transferred to a new 96-well PCR plate for individual purification with AMPure® XP beads (Beckman Coulter (UK) Ltd, UK) and an Invitrogen® magnetic stand (Fisher Scientific UK Ltd, UK). The Illumina PCR clean-up protocol was adapted to use 18.6 µL AMPure® XP beads (1.2x PCR product) to 15-16 µL PCR product. Illumina protocol was then followed until the beads were resuspended in 15 µL molecular grade water and incubated at room temperature for 5 minutes. The supernatant without beads in each well were not transferred to a new plate due to low volumes of purified product. Further pipetting may have resulted in loss of DNA. Each plate was sealed and stored at 4 °C until quality assurance. An Invitrogen™ Quant-IT™ PicoGreen™ dsDNA Assay (Fisher Scientific UK Ltd, UK) was conducted for all samples on a Fluoroskan™ Microplate Fluorometer (Life Technologies Ltd, UK). Samples were then normalised and pooled to create 4 nM pooled libraries before quantification using an Invitrogen™ Qubit™ dsDNA HS Assay Kit (Fisher Scientific UK Ltd, UK). Both libraries passed quality assurance with concentrations of 2.62 ng/µl and 4.14 ng/µl respectively. An Agilent 4200 Tapestation System (Agilent Technologies, CA, United States) was then used to check and compare size of the pooled libraries to selected samples. The pooled libraries were 272 bp and 299 bp (expected 286 bp) with samples in the same range. Equimolar libraries (4 nM) were then created using tapestation trace size estimates and Qubit concentrations. Libraries were run at 12 pM concentration on an Illumina MiSeq using 2 x 300 bp V3 chemistry (Illumina Inc., CA, USA). Both libraries included a 10% PhiX DNA spike-in control to improve clustering during initial sequencing.

Illumina data was converted from raw sequences to taxonomic assignment using a custom pipeline for reproducible analysis of metabarcoding data: metaBEAT (metaBarcoding and eDNA Analysis Tool) v0.8 (https://github.com/HullUni-bioinformatics/metaBEAT). Bioinformatic analysis using metaBEAT largely followed the workflow outlined by Hänfling *et al*. (2016) for sample processing and taxonomic assignment of sequenced eDNA samples from Windermere. Adaptations to this workflow are described (see also Harper *et al*. 2018a): raw reads were quality trimmed using Trimmomatic v0.32 (Bolger, Lohse, & Usadel, 2014), both from the read ends (minimum per base phred score Q30), as well as across sliding windows (window size 5bp; minimum average phred score Q30). Reads were clipped to a maximum length of 110 bp and reads shorter than 90 bp after quality trimming were discarded. To reliably exclude adapters and PCR primers, the first 25 bp of all remaining reads were also removed. Sequence

pairs were merged into single high quality reads using FLASH v1.2.11 (Magoč & Salzberg, 2011), if a minimum of 10 bp overlap with a maximum of 10% mismatch was detected between pairs. For reads that were not successfully merged, only forward reads were kept. To reflect our expectations with respect to fragment size, a final length filter was applied and only sequences of length 80-120 bp were retained. These were screened for chimeric sequences against our custom reference database using the uchime algorithm (Edgar *et al.*, 2011), as implemented in vsearch v1.1 (Rognes *et al.*, 2016). Redundant sequences were removed by clustering at 97% identity ('--cluster_fast' option) in vsearch v1.1 (Rognes *et al.*, 2016). Clusters represented by less than five sequences were considered sequencing error and omitted from further analyses. Non-redundant sets of query sequences were then compared against our custom reference database using BLAST (Zhang *et al.*, 2000). For any query matching with at least 98% identity to a reference sequence across more than 80% of its length, putative taxonomic identity was assigned using a lowest common ancestor (LCA) approach based on the top 10% BLAST matches. Sequences that could not be assigned (non-target sequences) were subjected to a separate BLAST search against the complete NCBI nucleotide (nt) database at 98% identity to determine the source via LCA as described above. To ensure reproducibility of analyses, the described workflow has been deposited in the GitHub repository for Chapter 2.

Data analysis

Manipulation of the metaBEAT dataset

Non-target sequence assignments and original assignments at 98% identity were merged. Any spurious assignments (i.e. non-UK species, invertebrates and bacteria) were removed from the dataset. Assignments to genera or families which contained only a single UK representative were manually assigned to that species. In our dataset, only genus *Strix* was reassigned to tawny owl (*Strix aluco*). Where family and genera assignments containing a single UK representative did have reads assigned to species, reads from all assignment levels were merged and manually assigned to that species. Consequently, all taxonomic assignments included in the final database were of species resolution. A total of 60 species were detected by eDNA metabarcoding. Mis-assignments in our dataset were then corrected; again, only one instance was identified. Scottish wildcat (*Felis*

silvestris) was reassigned to domestic cat (*Felis catus*) on the basis that Scottish wildcat does not occur where ponds were sampled (Kent, Lincolnshire and Cheshire).

GLMM comparison of eDNA methods for T. cristatus detection

Initially, a Poisson distribution was specified but tests using the R package RVAideMemoire v0.9-45-2 (Hervé, 2015) revealed models with this distribution were overdispersed. Models with a quasi-Poisson and zero-inflated distribution failed to resolve overdispersion (Ver Hoef & Boveng, 2007). A negative binomial distribution was used to control for aggregation in the count data and prevent biased parameter estimates (Harrison, 2014). Model overdispersion remained unresolved but model fit was improved. Model fit was assessed using the Hosmer and Lemeshow Goodness of Fit Test (Hosmer & Lemeshow, 2000) within the R package ResourceSelection v0.2-4 (Lele *et al.*, 2016). Model predictions were obtained using the *predictSE* function in the AICcmodavg package v2.0-3 (Mazerolle, 2017) and upper and lower 95% CIs were calculated from the standard error of the predictions.

Appendix 2.2: Supplementary results

In silico primer validation

The *in silico* analysis confirmed high taxonomic coverage (59.0% of target vertebrate species amplified) and resolution of the 12S rRNA primers. A wide range of UK vertebrate taxa were amplified, with fragment length ranging from 90-114 bp. The primers amplified 16/21 amphibian species, including *T. cristatus*. *L. helveticus*, Italian crested newt (*Triturus carnifex*), brown cave salamander (*Hydromantes genei*), edible frog (*Pelophylax esculentus*) and agile frog (*Rana dalmatina*) were not amplified *in silico*. All sequences from these species were manually aligned to the primers using the alignment viewer and editor AliView (Larsson, 2014), confirming potential for amplification. The primers amplified 47/67 fish species, including the threatened European eel (*Anguilla anguilla*), but amplification of UK freshwater fish assemblages was confirmed *in vitro* by Hänfling *et al*. (2016). The primers amplified 14/20 reptile species including slow worm (*Anguis fragilis*) and common lizard (*Zootoca vivipara*). Reference sequences were not available for one species and a further five species were not amplified. Primers were only validated for 282/621 bird species (including common waterfowl species). There were no 12S rRNA data available for 243/621 bird species and a further 96 species were not amplified. Similarly, no reference data were available for nine mammal species (bats and marine mammals) and a further 15 species were not amplified. Only 88/112 mammal species were validated. Several marine mammal species were not amplified but would not be found in freshwater ponds. However, priority species for freshwater management, such as European water vole (*Arvicola amphibious*) and American mink (*Mustela vison*), were not amplified alongside other species of bat, vole and shrew that may frequent ponds.

In vitro primer validation

Bands were observed by agarose gel electrophoresis for all amphibian tissue tested, including *L. helveticus* which was not amplified *in silico*, and no bands were observed in NTCs. The LOD was variable for each species: *T. cristatus*, *L. helveticus*, *R. temporaria* and *B. bufo* were not amplified below 5×10^{-4} ng/µl, whereas *I. alpestris* was not amplified below 5×10^{-3} ng/µl and *L. vulgaris* below 5×10^{-5} ng/µl. Due to sheer number

of and legislation surrounding many UK amphibian, reptile, bird, and mammal species, *in vitro* testing for all target taxa was unfeasible and metabarcoding proceeded on the basis of *in silico* amplification.

Appendix 2.3: Supplementary tables

Table S2.1: List of species for which no 12S rRNA records were available on GenBank. Only UK species which had no records for sister species within the same genus are included.

Common name	Binomial nomenclature
North Atlantic right whale	*Eubalaena glacialis*
Common kingfisher	*Alcedo atthis*
Trumpeter finch	*Bucanetes githagineus*
Green heron	*Butorides virescens*
Greater short-toed lark	*Calandrella brachydactyla*
Lesser short-toed lark	*Calandrella rufescens*
Lapland longspur	*Calcarius lapponicus*
Wilson’s warbler	*Cardellina pusilla*
Rufuous-tailed scrub robin	*Cercotrichas galactotes*
MacQueen’s bustard	*Chlamydotis macqueenii*
Lark sparrow	*Chondestes grammacus*
White-throated dipper	*Cinclus cinclus*
Great spotted cuckoo	*Clamator glandarius*
Long-tailed duck	*Clangula hyemalis*
Corn crake	*Crex crex*
Crested lark	*Galerida cristata*
European storm petrel	*Hydrobates pelagicus*
Little gull	*Hydrocoloeus minutus*
White-throated robin	*Irania gutturalis*
Hooded merganser	*Lophodytes cucullatus*
European crested tit	*Lophophanes cristatus*
Woodlark	*Lullula arborea*
Siberian blue robin	*Larvivora cyane*

Rufous-tailed robin	*Larvivora sibilans*
Thrush nightingale	*Luscinia luscinia*
Common nightingale	*Luscinia megarhynchos*
Bluethroat	*Luscinia svecica*
Black scoter	*Melanitta americana*
Velvet scoter	*Melanitta fusca*
Common scoter	*Melanitta nigra*
Surf scoter	*Melanitta perspicillata*
Bimaculated lark	*Melanocorypha bimaculata*
Calandra lark	*Melanocorypha calandra*
White-winged lark	*Melanocorypha leucoptera*
Black lark	*Melanocorypha yeltoniensis*
Song sparrow	*Melospiza melodia*
Black-and-white warbler	*Mniotilta varia*
Common rock thrush	*Monticola saxatilis*
Blue rock thrush	*Monticola solitarius*
Wilson's storm petrel	*Oceanites oceanicus*
Band-rumped storm petrel	*Oceanodroma castro*
Leach's storm petrel	*Oceanodroma leucorhoa*
Swinhoe's storm petrel	*Oceanodroma monorhis*
Tennessee warbler	*Oreothlypis peregrina*
Northern waterthrush	*Parkesia noveboracensis*
Savannah sparrow	*Passerculus sandwichensis*
Rosy starling	*Pastor roseus*
American cliff swallow	*Petrochelidon pyrrhonota*
Steller's eider	*Polysticta stelleri*
Eurasian crag martin	*Ptyonoprogne rupestris*
Sand martin	*Riparia riparia*
Whinchat	*Saxicola rubetra*

African stonechat	*Saxicola torquatus*
Northern parula	*Setophaga americana*
Hooded warbler	*Setophaga citrina*
American yellow warbler	*Setophaga petechia*
American redstart	*Setophaga ruticilla*
Wallcreeper	*Tichodroma muraria*
Brown thrasher	*Toxostoma rufum*
Golden-winged warbler	*Vermivora chrysoptera*

Table S2.2: List of species detected in PCR positive controls by eDNA metabarcoding and corresponding species-specific false positive sequence threshold applied.

Common name	Binomial name	False positive sequence threshold
European eel	*Anguilla anguilla*	0.000094
Common carp	*Cyprinus carpio*	0.000163
Common minnow	*Phoxinus phoxinus*	0.001287
Common roach	*Rutilus rutilus*	0.000291
European chub	*Squalius cephalus*	0.004080
Three-spined stickleback	*Gasterosteus aculeatus*	0.066667
Atlantic herring	*Clupea harengus*	0.000115
Common toad	*Bufo bufo*	0.066667
Common frog	*Rana temporaria*	0.000596
Smooth newt	*Lissotriton vulgaris*	0.066667
Great crested newt	*Triturus cristatus*	0.000276
Green-winged teal	*Anas carolinensis*	0.000322
Eurasian coot	*Fulica atra*	0.000223
Common moorhen	*Gallinula chloropus*	0.000179
Common starling	*Sturnus vulgaris*	0.000139
Human	*Homo sapiens*	0.253333
Brown rat	*Rattus norvegicus*	0.000467
Cow	*Bos taurus*	0.003542
Pig	*Sus scrofa*	0.000877

Table S2.3: Summary statistics for each Illumina MiSeq run.

MiSeq Run	Samples	Controls	Raw reads	Reads passing QC	Non-redundant reads	Reads taxonomically assigned	Unassigned reads
1	266	114	36,236,862	26,294,906	14,141,237	13,126,148	1,015,089
2	266	114	32,900,914	26,451,564	14,081,788	13,113,143	968,976

Table S2.4: Summary of read counts and the overall proportion of reads assigned to taxonomic levels for each Illumina MiSeq run.

MiSeq Run	Species	Genus	Family	Order	Class	Overall assignment (%)
1	10,185,014	1,438,216	963,865	12,454	526,599	92.82
2	9,419,096	1,237,427	1,899,932	10,723	545,965	93.12

Table S2.5: Summary of species detected by eDNA metabarcoding of freshwater pond samples ($N = 532$).

Common name	Binomial name	No. samples detected
European eel	*Anguilla anguilla*	15
Common barbel	*Barbus barbus*	2
Crucian carp	*Carassius carassius*	2
Common carp	*Cyprinus carpio*	41
Common minnow	*Phoxinus phoxinus*	13
Common roach	*Rutilus rutilus*	72
European chub	*Squalius cephalus*	21
Stone loach	*Barbatula barbatula*	15
Northern pike	*Esox lucius*	17
European bullhead	*Cottus gobio*	14
Three-spined stickleback	*Gasterosteus aculeatus*	56
Ninespine stickleback	*Pungitius pungitius*	15
Ruffe	*Gymnocephalus cernua*	1
Rainbow trout	*Oncorhynchus mykiss*	3
Common toad	*Bufo bufo*	42
Marsh frog	*Pelophylax ridibundus*	1
Common frog	*Rana temporaria*	120
Palmate newt	*Lissotrition helveticus*	5
Smooth newt	*Lissotriton vulgaris*	152
Great crested newt	*Triturus cristatus*	149
Green-winged teal	*Anas carolinensis*	7
Eurasian oystercatcher	*Haematopus ostralegus*	1
Common buzzard	*Buteo buteo*	4
Common pheasant	*Phasianus colchicus*	25
Domesticated turkey	*Meleagris gallopavo*	11

Helmeted guineafowl	*Numida meleagris*	1
Eurasian coot	*Fulica atra*	48
Common moorhen	*Gallinula chloropus*	215
Eurasian jay	*Garrulus glandarius*	7
European goldfinch	*Carduelis carduelis*	1
Dunnock	*Prunella modularis*	4
Eurasian nuthatch	*Sitta europaea*	1
Common starling	*Sturnus vulgaris*	4
Melodius warbler	*Hippolais polyglotta*	2
Grey heron	*Ardea cinerea*	1
Great spotted woodpecker	*Dendrocopus major*	1
Green woodpecker	*Picus viridis*	2
Tawny owl	*Strix aluco*	1
Dog	*Canis lupus*	65
Red fox	*Vulpes vulpes*	9
Eurasian otter	*Lutra lutra*	1
European badger	*Meles meles*	7
European polecat	*Mustela putorius*	1
Common pipistrelle	*Pipistrellus pipistrellus*	1
Eurasian water shrew	*Neomys fodiens*	9
Common shrew	*Sorex araneus*	1
European hare	*Lepus europaeus*	1
European rabbit	*Oryctolagus cuniculus*	24
Horse	*Equus caballus*	3
European water vole	*Arvicola amphibius*	16
Bank vole	*Myodes glareolus*	9
House mouse	*Mus musculus*	16
Brown rat	*Rattus norvegicus*	39
Grey squirrel	*Sciurus carolinensis*	57

Cow	*Bos taurus*	179
Sheep	*Ovis aries*	42
Red deer	*Cervus elaphus*	2
Reeve's muntjac	*Muntiacus reevesi*	3
Pig	*Sus scrofa*	140
Cat	*Felis catus*	16

Table S2.6: Summary of contaminants detected in PCR negative, or No Template Controls (NTCs), that occurred at high proportion of the total read count (> 1%). Maximum frequency and read count across all NTCs are provided for each contaminant.

Common name	Binomial name	No. NTCs detected	Max. proportion	Max. read count
Great crested newt	*Triturus cristatus*	6	93.0%	307
Smooth newt	*Lissotriton vulgaris*	12	100.0%	55
Common frog	*Rana temporaria*	10	63.2%	13,120
Common toad	*Bufo bufo*	1	22.8%	46
Common roach	*Rutilus rutilus*	6	81.3%	25,441
European bullhead	*Cottus gobio*	4	91.4%	10,827
Three-spined stickleback	*Gasterosteus aculeatus*	2	25.6%	166
Stone loach	*Barbatula barbatula*	1	6.2%	1,165
Common moorhen	*Gallinula chloropus*	4	41.8%	140
Mouse	*Mus musculus*	2	96.1%	1,759
Dog	*Canis lupus*	1	2.8%	18
Pig	*Sus scrofa*	1	97.7%	14,622
Sheep	*Ovis aries*	1	30.6%	589

Table S2.7: Summary of agreement (+) and disagreement (-) between egg searches, qPCR NT, qPCR TA, metabarcoding NT, and metabarcoding TA for *T. cristatus* detection in ponds (*N* = 532). NT represents No Threshold and TA represents Threshold Applied.

Method	Egg search +	qPCR NT +	qPCR TA +	Metabarcoding NT +	Metabarcoding TA +
Egg search -	*58 (+)* *448 (-)*	202	126	133	106
qPCR NT -	7	*265(+)* *267 (-)*	0	21	11
qPCR TA -	18	91	*174 (+)* *358 (-)*	48	26
Metabarcoding NT -	21	104	40	*182 (+)* *350 (-)*	0
Metabarcoding TA -	23	127	51	33	*149 (+)* *383 (-)*

Appendix 2.4: Supplementary figures

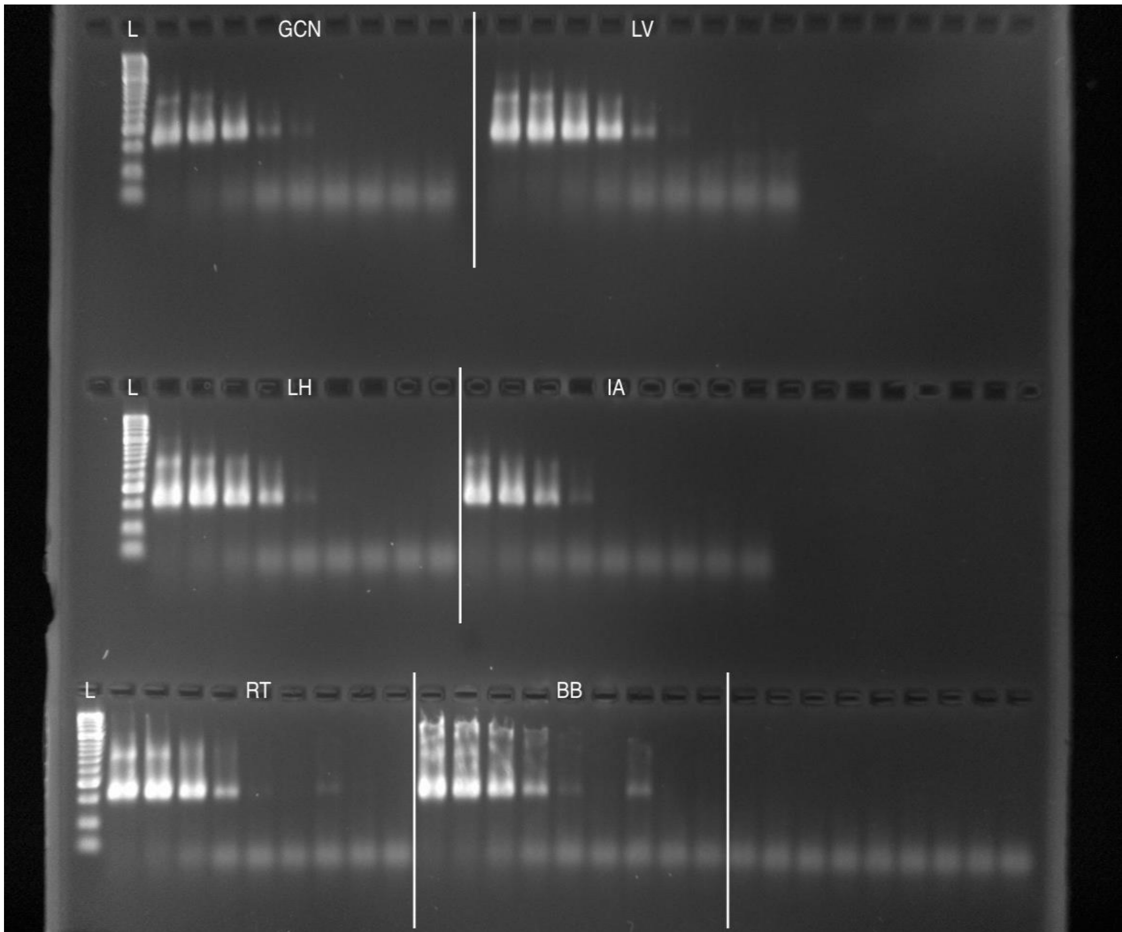


Figure S2.1: Gel image showing results of *in vitro* primer validation for each species: *T. cristatus* (GCN), *L. vulgaris* (LV), *L. helveticus* (LH), *I. alpestris* (IA), *R. temporaria* (RT) and *B. bufo* (BB).

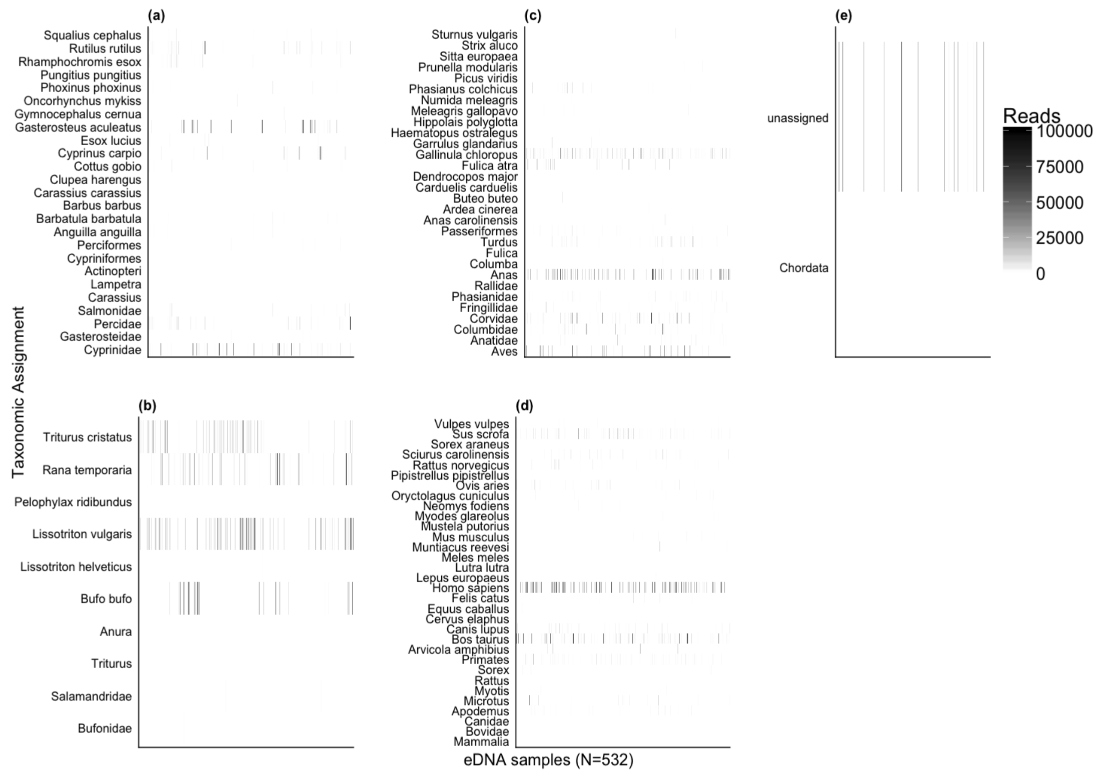


Figure S2.2: Heat maps of sequence read distribution for taxonomic assignments in each vertebrate group across all eDNA samples: **(a)** fish, **(b)** amphibians, **(c)** birds, **(d)** mammals and **(e)** other. Detections exceeding 100,000 reads (e.g. cow *Bos taurus*) were omitted during plotting to improve visualisation of lower read assignments in the dataset, but the data were not adjusted in this process. Each species was present in at least one sample although low read counts were not always visible.

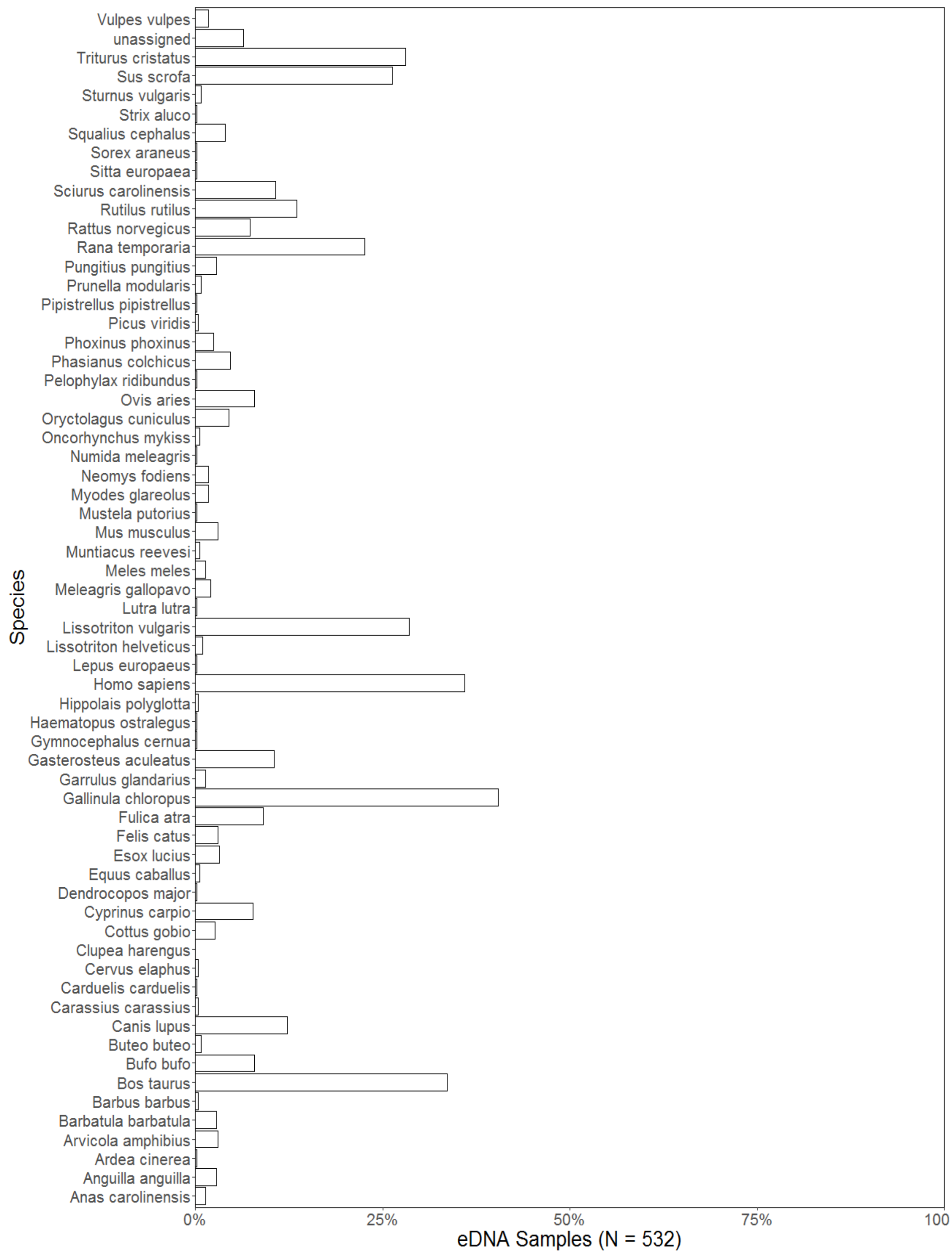


Figure S2.3: Proportion of eDNA samples in which each species was detected by eDNA metabarcoding.

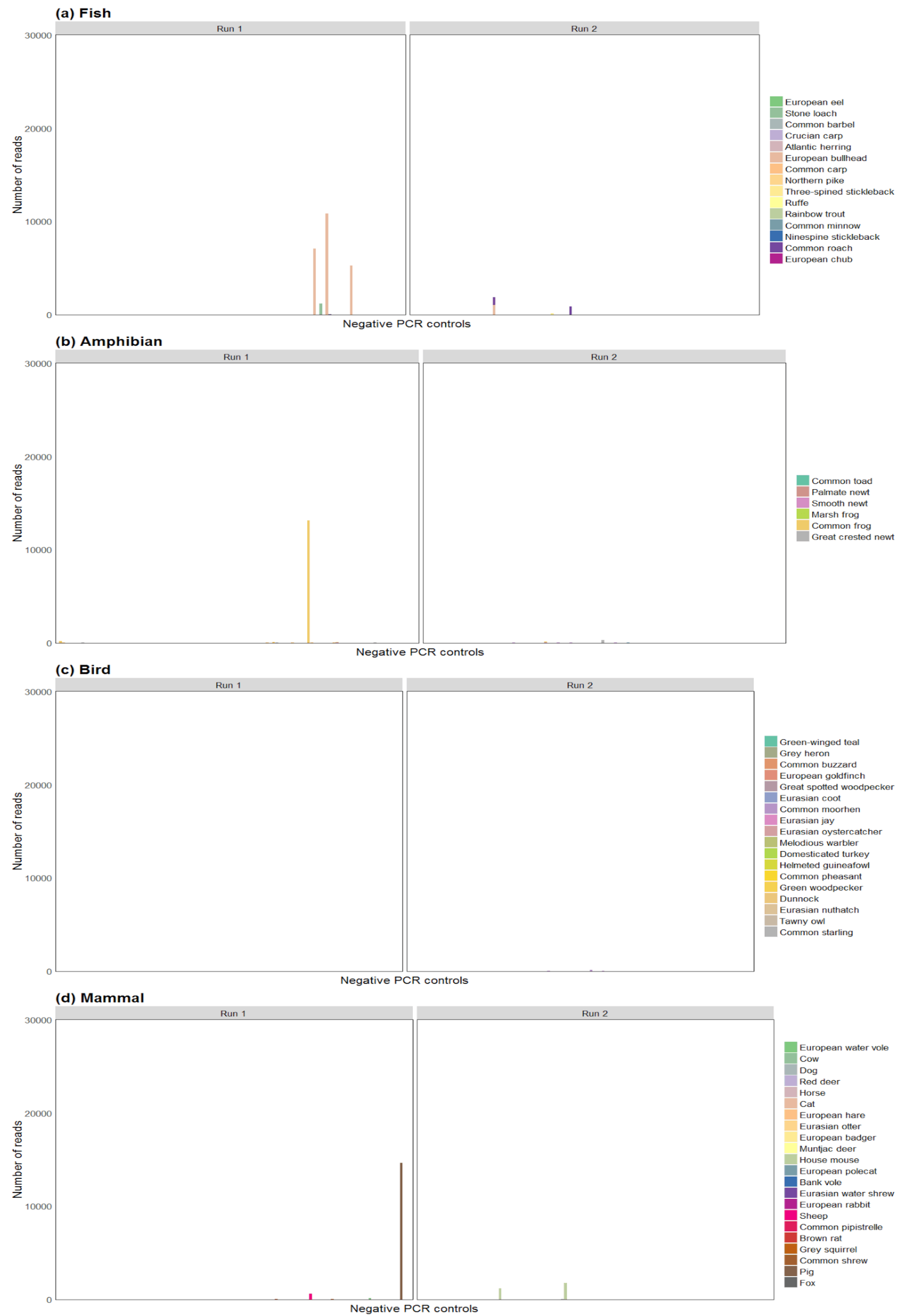


Figure S2.4: Presence of foreign DNA in PCR negative controls across sequencing runs. Highest contamination was observed from fish species, common roach (*Rutilus rutilus*) and European bullhead (*Cottus gobio*), in addition to *R. temporaria* and pig (*Sus scrofa*). *R. rutilus* occurred in six PCR negative controls, two of which exceeded 100 reads. *C. gobio* occurred in four PCR negative controls but all exceeded 1,000 reads. Notably, *R. temporaria* occurred in 13 PCR negative controls but only two exceeded 100 reads, with 180 and 13,120 reads. *S. scrofa* occurred in one PCR negative control only but exceeded 14,000 reads. Contamination from other species was relatively low with few species exceeding 100 sequence reads.

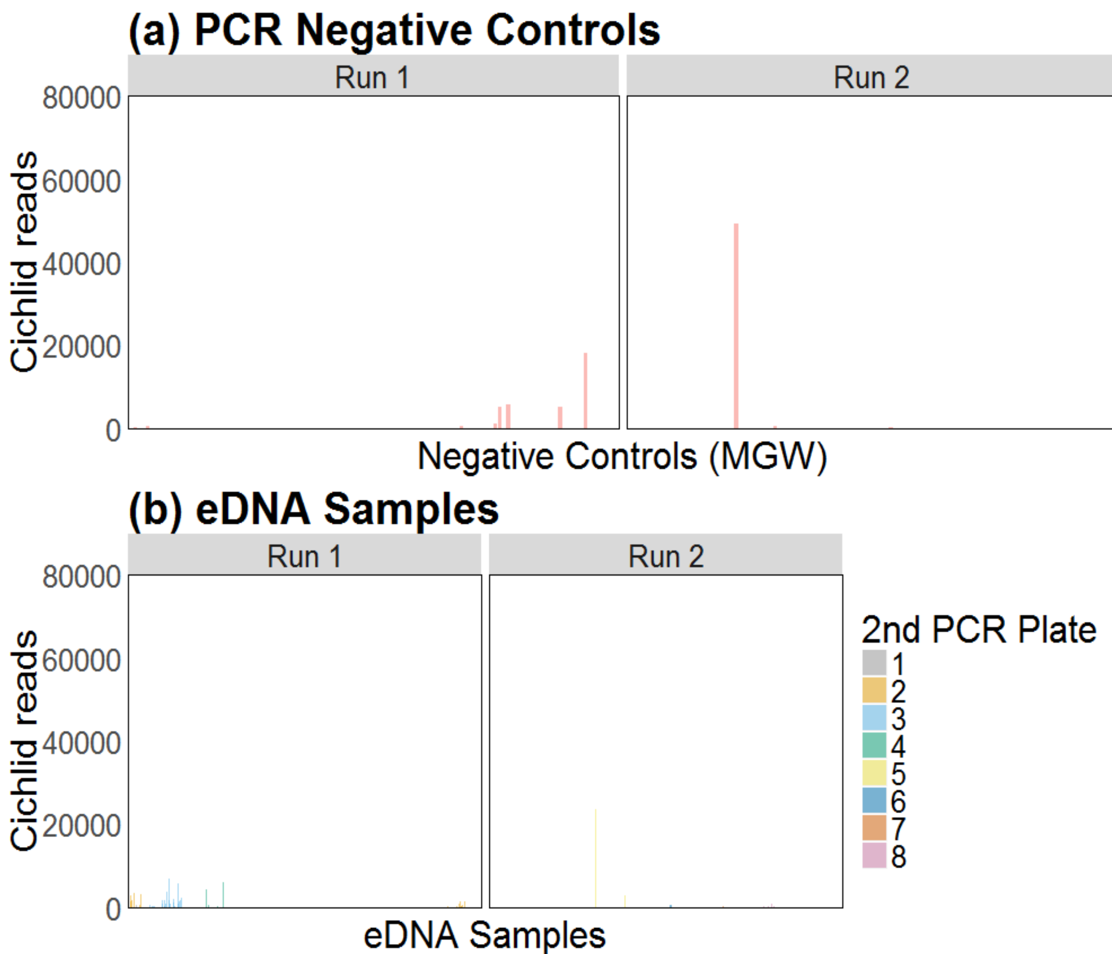


Figure S2.5: Presence of cichlid (*Rhamphochromis esox*) DNA (PCR positive control) amongst PCR negative controls and eDNA samples. Contamination of PCR negative controls was more frequent on the first sequencing run but greater where it occurred during the second sequencing run. Contamination of environmental samples was most common on plates 3 and 4, which were also sequenced on the first MiSeq run.

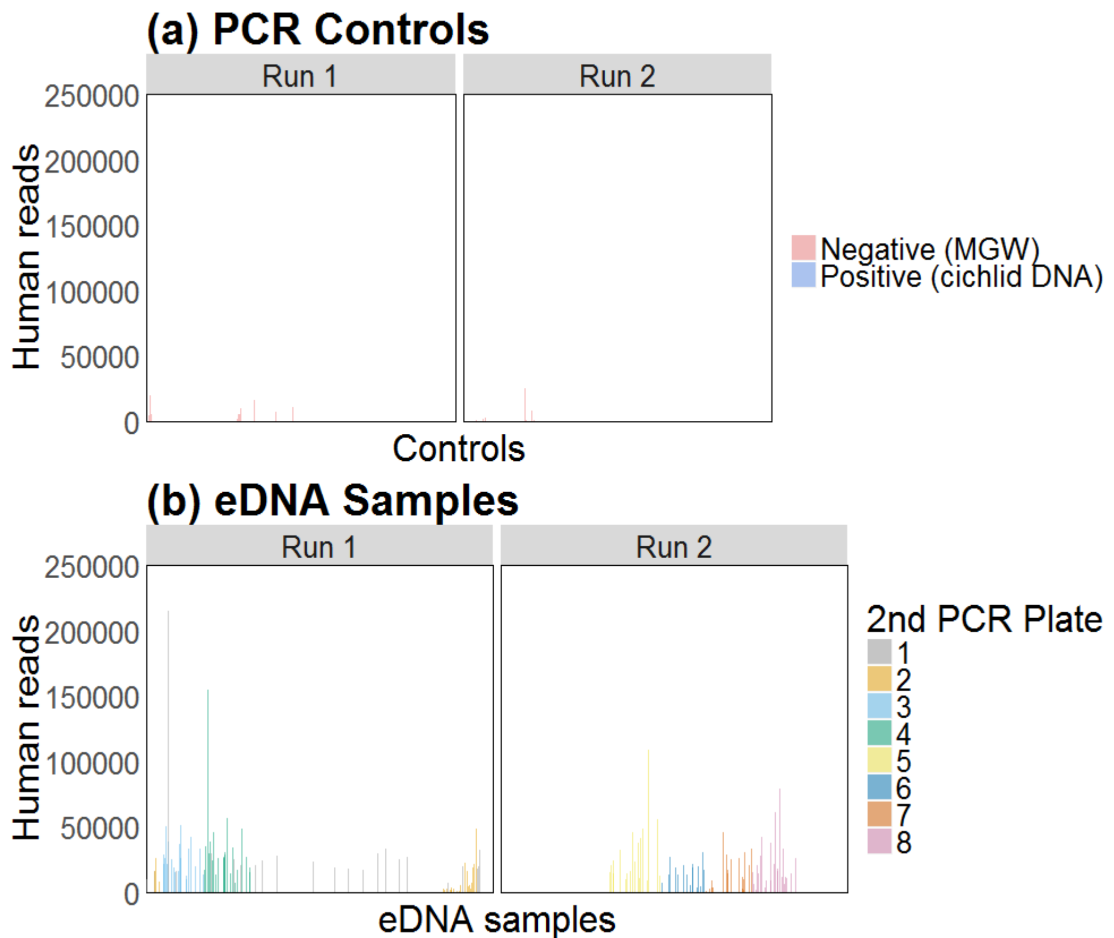


Figure S2.6: Presence of human (*Homo sapiens*) DNA amongst PCR controls and eDNA samples. Contamination of PCR controls and environmental samples was less frequent in the second sequencing run. *H. sapiens* DNA contamination was most abundant in environmental samples on PCR plates 1, 3, 4 and 5.

Appendix 2.5: Supplementary references

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.

Bolger, A. M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Ficetola, G.F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P. & Pompanon, F. (2010) An *In silico* approach for the evaluation of DNA barcodes. *BMC Genomics*, **11**, 434.

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.

Harper, L. R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H.C., Gough, K.C., Lewis, E., Adams, I.P., Brotherton, P., Phillips, S. & Hänfling, B. (2018) Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*). *Ecology and Evolution*, **8**, 6330–6341.

Harrison, X.A. (2014) Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ*, **2**, e616.

Hervé, M. (2015) RVAideMemoire: Diverse basic statistical and graphical functions. R package version 0.9-36.

Hosmer, D.W. & Lemeshow, S. (2000) Multiple Logistic Regression. *Applied Logistic Regression*, John Wiley & Sons, Inc., pp. 31–46.

Illumina (2011) *Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System*. Illumina Technical Note.

Kitson, J.J.N., Hahn, C., Sands, R.J., Straw, N.A., Evans, D.M. & Lunt, D.H. (2019) Detecting host-parasitoid interactions in an invasive Lepidopteran using nested tagging DNA-metabarcoding. *Molecular Ecology*, **28**, 471–483.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonier, M., Kyle, F.J.K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S. & Willing, C. (2016) Jupyter Notebooks – a publishing format for reproducible computational workflows. In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. IOS Press, pp. 87–90.

Kozlov, A.M., Zhang, J., Yilmaz, P., Glöckner, F.O. & Stamatakis, A. (2016) Phylogeny-aware identification and correction of taxonomically mislabeled sequences. *Nucleic Acids Research*, **44**, 5022-5033.

Larsson, A. (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, **30**, 3276–3278.

Lele, S.R., Keim, J.L. & Solymos, P. (2016) ResourceSelection: Resource Selection (Probability) Functions for Use-Availability Data. R package version 0.3-2.

Magoč, T. & Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.

Mazerolle, M.J. (2017) AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 2.1-1.

Natural History Museum (2017) UK Species Database. Accessed on 7th November 2017. http://www.nhm.ac.uk/our-science/data/uk-species/species/index.html

O'Donnell, J.L., Kelly, R.P., Lowell, N.C. & Port, J.A. (2016) Indexed PCR Primers Induce Template-Specific Bias in Large-Scale DNA Sequencing Studies. *PLoS One*, **11**, e0148698.

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.

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

Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, **4**, e2584.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Szitenberg, A., John, M., Blaxter, M.L. & Lunt, D.H. (2015) ReproPhylo: An Environment for Reproducible Phylogenomics. *PLoS Computational Biology*, **11**, e1004447.

Thomsen, P.F., 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.

Ver Hoef, J.M. & Boveng, P.L. (2007) QUASI-POISSON VS. NEGATIVE BINOMIAL REGRESSION: HOW SHOULD WE MODEL OVERDISPERSED COUNT DATA? *Ecology,* **88**, 2766-2772.

Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, **7**, 203–214.

Appendix 3

Appendix 3.1: Supplementary methods

Data analysis

A supplementary analysis was performed where a series of blanket false positive sequence thresholds (0.05 - 30%) were applied to the dataset to ensure results did not differ drastically from species-specific thresholds (see Tables S3.4-3.9).

Individual species associations

Species associations between all vertebrates were investigated using presence-absence data generated by eDNA metabarcoding with the method of Veech (2013) implemented in the R package cooccur v1.3 (Griffith, Veech & Marsh, 2016). This is a probabilistic model which measures species co-occurrence (presence-absence) as the number of sampling sites where two species co-occur. The observed co-occurrence of a given dataset is compared to the expected co-occurrence. Expected co-occurrence is determined by the probabilities of each species' occurrence multiplied by the number of sampling sites. Effect sizes were also computed for species pairs to examine species associations regardless of statistical significance. These are equivalent to the difference between expected and observed frequency of co-occurrence. The values are then standardized by dividing these differences by sample size. In standardized form, these values are bounded from -1 to 1, with positive values indicating positive associations and negative values indicating negative associations.

Biotic and abiotic determinants of T. cristatus occurrence

Collinearity and spatial autocorrelation within the dataset were investigated before the most appropriate regression model was determined. Collinearity between explanatory variables was assessed using a Spearman's rank pairwise correlation matrix. Collinearity was observed between pond circumference, pond length, pond width, and pond area. Pond area encompasses length and width thus taking the same measurements and accounting for the same variance in the data as these variables. Therefore, pond circumference, pond

length, and pond width were removed from the dataset so as remaining variables were not highly correlated (Zuur *et al*., 2009). Shading (percentage of total pond margin shaded) and terrestrial overhang (percentage of pond overhung by trees and shrubs) were also collinear. As terrestrial overhang accounts for shading of the entire pond, whereas shading considers only the pond margin, terrestrial overhang was retained as an explanatory variable. After collinear variables were removed, variance inflation factors (VIFs) of remaining variables were calculated using the R package car v2.1-6 (Fox & Weisberg, 2011) to identify remnant multicollinearity. Multicollinearity (VIF > 3) (Zuur *et al*., 2009) was still present in Habitat Suitability Index (HSI) score and HSI band. Many of the environmental variables are also used as indices to calculate HSI score thus HSI score may mask variation caused by these variables individually. HSI score and HSI band were removed prior to model selection.

A large number of explanatory variables remained: max. depth; area; density, overhang; macrophyte cover; permanence; water quality; pond substrate; inflow; outflow; pollution; presence of amphibians, waterfowl and fish; woodland; rough grass; scrub/hedge; ruderals; terrestrial other; and overall terrestrial habitat quality. The relative importance of these for determining *T. cristatus* occurrence was inferred using a classification tree within the R package rpart v4.1-13 (Therneau, Atkinson & Ripley, 2014). The classification tree suggested the most important explanatory variables of *T. cristatus* occurrence were: *L. vulgaris* presence, species richness, maximum depth of ponds, fish presence, pond density, pond area, amphibian presence, waterfowl presence (which incorporates identified species associations between *T. cristatus* and common moorhen [*Gallinula chloropus*] and Eurasian coot [*Fulica atra*]), terrestrial habitat, pond substrate, grey squirrel (*Sciurus carolinensis*) presence, three-spined stickleback presence (*Gasterosteus aculeatus*), pond outflow, macrophyte cover, water quality and pond permanence. *L. vulgaris*, *S. carolinensis*, and *G. aculeatus* were also identified as having significant associations with *T. cristatus* by the co-occurrence analysis. A pruning diagram was applied to the data to cross-validate the classification tree and remove unimportant explanatory variables. A tree of six was optimal according to the pruning diagram, indicating that six explanatory variables should be retained for statistical analysis. Many variables occurred more than once in the classification tree, indicative of weak non-linear relationships with the response variable. Generalized Additive Models (GAMs) were performed to deal with non-linearity but several explanatory variables were in fact linear, i.e. estimated one degree of freedom for smoother (Zuur *et al*., 2009).

The ponds in this study had restricted spatial distribution and were nested within three UK counties (Figure S3.1) thus spatial autocorrelation may be present. This phenomena is common in ecological studies of species presence-absence as sites located within an animal's ranging capability are likely to be inhabited (Zuur *et al*., 2009). *T. cristatus* individuals can migrate distances of 1-2 km to new ponds (Edgar & Bird, 2006; Haubrock & Altrichter, 2016), thus occurrence of *T. cristatus* is likely in ponds that are closely located to one another in a given area. Spline correlograms - graphical representations of spatial correlation between locations at a range of lag distances that are smoothed using a spline function (Bjørnstad, 2017) - were constructed using R package ncf v1.1-7 to examine spatial autocorrelation between ponds. Spline correlograms of the Pearson residuals of the raw data, a binomial Generalized Linear Model (GLM), and a binomial Generalized Linear Mixed Model (GLMM) were compared. GLMMs can account for dependencies within sites, handled with the introduction of random effects (Zuur *et al*., 2009). Each eDNA sample represented a different pond and thus sample was treated as a random effect. The GLMM successfully accounted for spatial dependencies between ponds based on the spline correlogram of the Pearson residuals. A series of alternative mixed effects models that covered different combinations of explanatory variables to test different hypotheses were then evaluated. Explanatory variables were grouped into functional groups. For example, pond properties, terrestrial habitat and pond biodiversity. The GLMM containing only presence of species or guilds had the lowest AIC value but as we were also interested in habitat predictors of *T. cristatus*, model selection was performed on the GLMM containing all explanatory variables.

Biotic and abiotic determinants of vertebrate species richness

The species richness classification tree indicated that terrestrial overhang was the most important explanatory variable, followed by amphibian presence, rough grass habitat, pond density, maximum pond depth, pond area, woodland, ruderals, pollution, fish presence, terrestrial other, macrophyte cover, pond outflow, water quality, waterfowl presence, pond inflow, scrub/hedge and pond permanence. A tree of three or five was optimal according to the pruning diagram, indicating that three or five explanatory variables should be retained for statistical analysis.

Appendix 3.2: Supplementary results

Biotic and abiotic determinants of *T. cristatus* occurrence

The co-occurrence analysis revealed of 1770 species pair combinations. 1406 pairs (79.44%) were removed from the analysis because expected co-occurrence was less than one, leaving 364 pairs for analysis. The pairwise combinations revealed 17 negative and 48 positive significant co-occurrence patterns. The remaining co-occurrence patterns were random thus the observed presence-absence data did not significantly deviate from the expected presence-absence data. No pairs were unclassifiable indicative of sufficient statistical power to analyse all pairs. A pairing profile was constructed to understand each species' individual contribution to the positive and negative species associations. Interactions were clustered in a few species rather than being evenly distributed. When observed and expected co-occurrence was examined, some species pairs deviated from the expected co-occurrence. A minority of species pairs exhibited fewer than expected co-occurrences but these pairs were largely clustered towards having low expected co-occurrence.

Appendix 3.3: Supplementary tables

Table S3.1: Summary of environmental metadata on pond characteristics and surrounding terrestrial habitat included in analysis of *T. cristatus* occupancy and vertebrate species richness.

Variable	Description	Unit/categories
Maximum depth	Depth of pond	m
Circumference	Pond circumference	m
Width	Pond width	m
Length	Pond length	m
Area	Pond area	m^2
Density	Pond density	Number of ponds per km^2
Terrestrial overhang	Percentage of pond overhung by trees and shrubs	%
Shading	Percentage of total pond margin shaded to at least 1 m from the shore	%
Macrophyte cover	Percentage of pond surface occupied by macrophytes	%
Habitat Suitability Index (HSI)	Score calculated from aforementioned variables which indicates habitat quality for crested newt (0 = poor, 1 = excellent)	Decimal
Habitat Suitability Index (HSI) band	Categorical classification of HSI score	Poor/below average/average/good
Pond permanence	Pond permanence	Dries annually/rarely dries/sometimes dries/ never dries
Water quality	Subjective assessment based on invertebrate diversity, presence of submerged vegetation, and	Bad/poor/moderate/goo d/excellent

	knowledge of water inputs to pond.	
Pond substrate	Type of substrate	Not known/rock/clay/concrete/sand, gravel, pebbles/lined/peat-organic
Inflow	Water inputs to pond	Absent/present
Outflow	Water leaving pond	Absent/present
Pollution	Rubbish or other signs of pollution	Absent/present
Other amphibians	Presence of amphibian species other than crested newt	Absent/present
Fish	Presence of any fish species	Absent/possible/minor/ major
Waterfowl	Presence of any waterfowl species	Absent/minor/major
Woodland	Terrestrial habitat: woodland	None/some/important
Rough grass	Terrestrial habitat: rough grass	None/some/important
Scrub/hedge	Terrestrial habitat: scrub/hedge	None/some/important
Ruderals	Terrestrial habitat: ruderals	None/some/important
Terrestrial other	Other good quality terrestrial habitat that does not conform to aforementioned habitat types	None/some/important
Overall terrestrial habitat score	Overall quality of terrestrial habitat	None/poor/moderate/good

Table S3.2: List of species for which no 12S rRNA records were available on GenBank. Only UK species which had no records for sister species within the same genus are included.

Common name	Binomial nomenclature
North Atlantic right whale	*Eubalaena glacialis*
Common kingfisher	*Alcedo atthis*
Trumpeter finch	*Bucanetes githagineus*
Green heron	*Butorides virescens*
Greater short-toed lark	*Calandrella brachydactyla*
Lesser short-toed lark	*Calandrella rufescens*
Lapland longspur	*Calcarius lapponicus*
Wilson's warbler	*Cardellina pusilla*
Rufuous-tailed scrub robin	*Cercotrichas galactotes*
MacQueen's bustard	*Chlamydotis macqueenii*
Lark sparrow	*Chondestes grammacus*
White-throated dipper	*Cinclus cinclus*
Great spotted cuckoo	*Clamator glandarius*
Long-tailed duck	*Clangula hyemalis*
Corn crake	*Crex crex*
Crested lark	*Galerida cristata*
European storm petrel	*Hydrobates pelagicus*
Little gull	*Hydrocoloeus minutus*
White-throated robin	*Irania gutturalis*
Hooded merganser	*Lophodytes cucullatus*
European crested tit	*Lophophanes cristatus*
Woodlark	*Lullula arborea*
Siberian blue robin	*Larvivora cyane*
Rufous-tailed robin	*Larvivora sibilans*
Thrush nightingale	*Luscinia luscinia*
Common nightingale	*Luscinia megarhynchos*
Bluethroat	*Luscinia svecica*
Black scoter	*Melanitta americana*

Velvet scoter	*Melanitta fusca*
Common scoter	*Melanitta nigra*
Surf scoter	*Melanitta perspicillata*
Bimaculated lark	*Melanocorypha bimaculata*
Calandra lark	*Melanocorypha calandra*
White-winged lark	*Melanocorypha leucoptera*
Black lark	*Melanocorypha yeltoniensis*
Song sparrow	*Melospiza melodia*
Black-and-white warbler	*Mniotilta varia*
Common rock thrush	*Monticola saxatilis*
Blue rock thrush	*Monticola solitarius*
Wilson’s storm petrel	*Oceanites oceanicus*
Band-rumped storm petrel	*Oceanodroma castro*
Leach’s storm petrel	*Oceanodroma leucorhoa*
Swinhoe’s storm petrel	*Oceanodroma monorhis*
Tennessee warbler	*Oreothlypis peregrina*
Northern waterthrush	*Parkesia noveboracensis*
Savannah sparrow	*Passerculus sandwichensis*
Rosy starling	*Pastor roseus*
American cliff swallow	*Petrochelidon pyrrhonota*
Steller’s eider	*Polysticta stelleri*
Eurasian crag martin	*Ptyonoprogne rupestris*
Sand martin	*Riparia riparia*
Whinchat	*Saxicola rubetra*
African stonechat	*Saxicola torquatus*
Northern parula	*Setophaga americana*
Hooded warbler	*Setophaga citrina*
American yellow warbler	*Setophaga petechia*
American redstart	*Setophaga ruticilla*
Wallcreeper	*Tichodroma muraria*
Brown thrasher	*Toxostoma rufum*
Golden-winged warbler	*Vermivora chrysoptera*

Table S3.3: List of species detected in PCR positive controls by eDNA metabarcoding and corresponding species-specific false positive sequence threshold applied.

Common name	Binomial name	Species-specific false positive sequence threshold
European eel	*Anguilla anguilla*	0.000094
Common carp	*Cyprinus carpio*	0.000163
Common minnow	*Phoxinus phoxinus*	0.001287
Common roach	*Rutilus rutilus*	0.000291
European chub	*Squalius cephalus*	0.004080
Three-spined stickleback	*Gasterosteus aculeatus*	0.066667
Atlantic herring	*Clupea harengus*	0.000115
Common toad	*Bufo bufo*	0.066667
Common frog	*Rana temporaria*	0.000596
Smooth newt	*Lissotriton vulgaris*	0.066667
Great crested newt	*Triturus cristatus*	0.000276
Green-winged teal	*Anas carolinensis*	0.000322
Eurasian coot	*Fulica atra*	0.000223
Common moorhen	*Gallinula chloropus*	0.000179
Common starling	*Sturnus vulgaris*	0.000139
Human	*Homo sapiens*	0.253333
Brown rat	*Rattus norvegicus*	0.000467
Cow	*Bos taurus*	0.003542
Pig	*Sus scrofa domesticus*	0.000877

Table S3.4: Effect of number of species in different vertebrate groups on *T. cristatus* occupancy as determined using a binomial GLMM for different metabarcoding sequence thresholds ($N = 532$ ponds). For categorical variables with more than one level, effect size and standard error are only given for levels reported in the model summary. Test statistic is for LRT used. Significant P-values (<0.05) are in bold.

Threshold	Overdispersion	Model fit	Model variables	Effect size	Standard error	χ^2	*P*
No threshold	$\chi^2_{525} = 519.016$ $P = 0.566$	$\chi^2_8 = 18.319$ $P = 0.019$ $R^2 = 10.10\%$	**Fish**	-0.215	0.101	4.913	**0.027**
			Amphibian	0.454	0.120	16.528	**<0.001**
			Waterfowl	0.523	0.163	11.070	**0.001**
			Terrestrial bird	-0.435	0.277	2.715	0.099
			Mammal	0.146	0.082	3.224	0.073
0.05%	$\chi^2_{525} = 526.993$ $P = 0.467$	$\chi^2_8 = 56.79$ $P < 0.001$ $R^2 = 6.93\%$	**Fish**	-0.238	0.121	4.224	**0.040**
			Amphibian	0.338	0.127	7.723	**0.006**
			Waterfowl	0.547	0.178	10.163	**0.001**
			Terrestrial bird	-0.399	0.315	1.786	0.182
			Mammal	-0.007	0.089	0.005	0.941
0.1%	$\chi^2_{525} = 526.839$ $P = 0.469$	$\chi^2_8 = 17.728$ $P = 0.023$ $R^2 = 7.03\%$	**Fish**	-0.241	0.130	3.781	0.052
			Amphibian	0.360	0.130	8.471	**0.004**
			Waterfowl	0.544	0.180	9.813	**0.002**
			Terrestrial bird	-0.356	0.315	1.401	0.237
			Mammal	-0.036	0.092	0.157	0.692
0.5%	$\chi^2_{525} = 539.371$ $P = 0.323$	$\chi^2_8 = 9.141$ $P = 0.331$ $R^2 = 9.91\%$	**Fish**	-0.331	0.155	5.150	**0.023**
			Amphibian	0.328	0.132	6.177	**0.013**
			Waterfowl	0.633	0.180	12.400	**<0.001**
			Terrestrial bird	-0.962	0.465	5.714	**0.017**
			Mammal	0.067	0.108	0.380	0.538
1%	$\chi^2_{525} = 515.411$ $P = 0.609$	$\chi^2_8 = 15.946$ $P = 0.043$ $R^2 = 14.45\%$	**Fish**	-0.547	0.206	9.077	**0.003**
			Amphibian	0.405	0.153	8.260	**0.004**
			Waterfowl	0.654	0.210	11.246	**0.001**
			Terrestrial bird	-1.639	0.736	9.060	**0.003**
			Mammal	0.047	0.130	0.133	0.716
5%	Model could not be fit to the data.						
10%	$\chi^2_{525} = 0.405$ $P = 1.000$	$\chi^2_8 = 0.382$ $P = 1.000$ $R^2 = 98.83\%$	**Fish**	-0.023	52.42	0.398	0.528
			Amphibian	0.039	11.63	162.241	**<0.001**
			Waterfowl	0.091	15.65	0.920	0.338
			Terrestrial bird	3.97×10^3	2.54×10^7	3.559	0.059

			Mammal	-0.049	19.67	7.150	**0.008**
30%	Model could not be fit to the data.						
Species-specific	$\chi^2_{525} = 517.497$ $P = 0.584$	$\chi^2_8 = 22.581$ $P = 0.004$ $R^2 = 9.41\%$	**Fish**	-0.238	0.124	4.049	**0.044**
			Amphibian	0.557	0.149	16.564	**<0.001**
			Waterfowl	0.621	0.181	13.229	**<0.001**
			Terrestrial bird	-0.328	0.291	1.383	0.240
			Mammal	0.016	0.090	0.032	0.858

Table S3.5: Summary of different significant associations between *T. cristatus* and other vertebrate species as determined by the probabilistic co-occurrence model at different metabarcoding sequence thresholds (*N* = 532 ponds).

Threshold	Positive pairs	Negative pairs	Random pairs	Positive associations with *T. cristatus*		Negative associations with *T. cristatus*	
				Species	***P***	**Species**	***P***
None	64	4	338	*Bos taurus*	<0.001	*Cyprinus carpio*	0.029
				Fulica atra	0.007		
				Gallinula chloropus	<0.001		
				Lissotriton vulgaris	<0.001		
				Sus scrofa domesticus	<0.001		
0.05%	53	6	296	*Fulica atra*	0.027	*Bufo bufo*	0.003
				Gallinula chloropus	<0.001	*Gasterosteus aculeatus*	0.003
				Lissotriton vulgaris	<0.001	*Sciurus carolinensis*	0.032
				Sus scrofa domesticus	0.002		
0.1%	47	7	277	*Fulica atra*	0.032	*Bufo bufo*	0.011
				Gallinula chloropus	0.001	*Gasterosteus aculeatus*	0.009
				Lissotriton vulgaris	<0.001	*Sciurus carolinensis*	0.023
				Sus scrofa domesticus	0.009		
0.5%	37	13	205	*Fulica atra*	0.008	*Bufo bufo*	0.006
				Gallinula chloropus	0.001	*Gasterosteus aculeatus*	0.009
				Lissotriton vulgaris	<0.001	*Sciurus carolinensis*	0.005
				Sus scrofa domesticus	0.004	*Esox Lucius*	0.031
						Phasianus colchicus	0.023
1%	23	9	169	*Gallinula chloropus*	0.001	*Bufo bufo*	0.010
				Lissotriton vulgaris	<0.001	*Gasterosteus aculeatus*	0.001
				Sus scrofa domesticus	0.014	*Sciurus carolinensis*	0.042
						Esox Lucius	0.044
						Phasianus colchicus	0.012
5%	3	7	76	*Gallinula chloropus*	0.007	*Bufo bufo*	0.004
				Lissotriton vulgaris	<0.001	*Gasterosteus aculeatus*	0.004
						Cyprinus carpio	0.029
10%	2	3	51	*Lissotriton vulgaris*	<0.001	*Bufo bufo*	0.020
						Gasterosteus aculeatus	0.003
30%	0	1	11				
Species-specific	48	17	299	*Fulica atra*	0.023	*Bufo bufo*	0.009
				Gallinula chloropus	0.001	*Gasterosteus aculeatus*	0.009
				Lissotriton vulgaris	< 0.001	*Sciurus carolinensis*	0.018

	Sus scrofa domesticus	0.004	*Phasianus colchicus*	0.048
			Pungitius pungitius	0.047

Table S3.6: Summary of abiotic and biotic determinants of *T. cristatus* occupancy as identified using a binomial GLMM for different metabarcoding sequence thresholds (n = 504 ponds). For categorical variables with more than one level, effect size and standard error are only given for levels reported in the model summary. Test statistic is for LRT used. Significant P-values (<0.05) are in bold.

Threshold	Overdispersion	Model fit	Model variables	Effect size	Standard error	χ^2	*P*
No threshold	χ^2_{496} = 525.999	χ^2_8 = 14.167	***L. vulgaris***	1.303	0.252	29.174	**<0.001**
	P = 0.170	P = 0.078	**Species richness**	0.305	0.053	37.618	**<0.001**
		R^2 = 33.94%	**Inflow**	-0.757	0.244	10.029	**0.002**
			Ruderals			6.690	**0.035**
			None	-0.813	0.455		
			Some	-0.313	0.466		
			C. carpio	-1.584	0.501	12.374	**<0.001**
0.05%	χ^2_{490} = 405.328	χ^2_8 = 6.171	***L. vulgaris***	0.635	0.278	5.794	**0.016**
	P = 0.998	P = 0.628	**Species richness**	0.510	0.104	52.263	**<0.001**
		R^2 = 40.99%	***B. bufo***	-1.936	0.505	24.704	**<0.001**
			S. carolinensis	-2.140	0.603	19.946	**<0.001**
			G. aculeatus	-1.703	0.503	17.317	**<0.001**
			Inflow	-0.913	0.306	10.671	**0.001**
			Pond area	0.0004	0.0002	5.726	**0.017**
			Permanence	0.482	0.492	7.934	**0.047**
			Never dries				
			Rarely dries	0.213	0.539		
			Sometimes dries	-0.420	0.530		
			Ruderals			6.055	**0.048**
			None	-0.567	0.552		
			Some	0.067	0.551		
0.1%	χ^2_{488} = 407.611	χ^2_8 = 6.232	**Species richness**	0.510	0.115	82.906	**< 0.001**
	P = 0.997	P = 0.621	***B. bufo***	-1.844	0.518	21.710	**<0.001**
		R^2 = 41.00%	**Inflow**	-0.866	0.311	9.350	**0.002**
			S. carolinensis	-2.386	0.666	20.517	**<0.001**
			Max. depth	0.403	0.143	9.144	**0.003**
			G. aculeatus	-1.623	0.495	16.589	**<0.001**
			Macrophytes	0.010	0.005	4.493	**0.034**
			Pond area	0.0005	0.0002	7.730	**0.005**
			Ruderals			9.752	**0.008**
			None	-0.698	0.542		
			Some	0.107	0.543		
			Woodland			7.375	**0.025**
			None	-0.874	0.366		
			Some	-0.279	0.322		
			Terrestrial other			7.324	**0.026**
			None	0.322	0.456		
			Some	-0.402	0.446		

0.5%	$\chi^2_{491} = 352.876$	$\chi^2_{8} = 17.172$	**Species richness**	0.739	0.158	83.028	**<0.001**
	$P = 0.999$	$P = 0.028$	***B. bufo***	-2.227	0.641	23.505	**<0.001**
		$R^2 = 47.27\%$	**Inflow**	-1.421	0.402	21.583	**<0.001**
			Pond area	0.0006	0.0003	6.955	**0.008**
			G. aculeatus	-1.847	0.588	15.679	**<0.001**
			Permanence			18.733	**<0.001**
			Never dries	0.950	0.543		
			Rarely dries	0.689	0.576		
			Sometimes dries	-0.595	0.574		
			S. carolinensis	-3.126	0.881	26.827	**<0.001**
			Woodland			9.606	**0.008**
			None	-0.961	0.401		
			Some	-0.143	0.340		
1%	$\chi^2_{496} = 485.663$	$\chi^2_{8} = 5.940$	**Species richness**	0.608	0.130	56.081	**<0.001**
	$P = 0.622$	$P = 0.654$	**Overhang**	-0.011	0.004	8.463	**0.004**
		$R^2 = 38.34\%$	***G. aculeatus***	-2.132	0.632	20.225	**<0.001**
			Pond area	0.0006	0.0002	10.201	**0.001**
			Inflow	-1.144	0.340	16.056	**<0.001**
			Max. depth	0.266	0.134	4.319	**0.038**
5%	Model could not be fit to the data.						
10%	No explanatory variables retained by model selection - null model had better fit than final model from model selection. Due to threshold stringency and highly reduced detection of great crested newt, no explanatory variables adequately fit the data.						
30%	No explanatory variables retained by model selection - null model had better fit than final model from model selection. Due to threshold stringency and highly reduced detection of great crested newt, no explanatory variables adequately fit the data.						
Species-specific	$\chi^2_{496} = 485.663$	$\chi^2_{8} = 5.940$	***L. vulgaris***	1.081	0.303	17.434	**<0.001**
	$P = 0.622$	$P = 0.6540$	**Species richness**	0.527	0.105	60.267	**<0.001**
		$R^2 = 38.34\%$	***B. bufo***	-1.635	0.696	8.228	**0.004**
			S. carolinensis	-1.591	0.534	12.432	**<0.001**
			G. aculeatus	-1.432	0.561	9.453	**0.002**
			Pond area	0.0004	0.0002	6.453	**0.011**
			Pond depth	0.282	0.139	4.266	**0.039**
			Outflow	-0.713	0.359	4.467	**0.035**
			Ruderals			6.507	**0.039**
			None	-0.617	0.527		
			Some	0.032	0.528		
			Terrestrial other			7.918	**0.019**
			None	0.428	0.429		
			Some	-0.316	0.424		

Table S3.7: Summary of relationship between HSI score and *T. cristatus* occupancy as determined using a binomial GLMM for different metabarcoding sequence thresholds (n = 504 ponds). Test statistic is for LRT used. Significant P-values (<0.05) are in bold.

Threshold	GLMM results	Overdispersion	Model fit
None	2.649 ± 0.735 χ^2_1 = 13.791 ***P* < 0.001**	χ^2_{501} = 506.140 P = 0.428	χ^2_8 = 4.801 P = 0.779 R^2 = 3.88%
0.05%	3.070 ± 0.795 χ^2_1 = 16.114 ***P* < 0.001**	χ^2_{501} = 507.131 P = 0.415	χ^2_8 = 8.880 P = 0.353 R^2 = 5.14%
0.1%	3.081 ± 0.805 χ^2_1 = 15.831 ***P* < 0.001**	χ^2_{501} = 507.366 P = 0.412	χ^2_8 = 9.902 P = 0.272 R^2 = 5.18%
0.5%	3.3863 ± 0.841 χ^2_1 = 17.739 ***P* < 0.001**	χ^2_{501} = 510.637 P = 0.373	χ^2_8 = 14.558 P = 0.068 R^2 = 6.19%
1%	3.775 ± 0.887 χ^2_1 = 20.163 ***P* < 0.001**	χ^2_{501} = 511.628 P = 0.362	χ^2_8 = 16.657 ***P* = 0.034** R^2 = 7.58%
5%	Null model better fit to data. *T. cristatus* occupancy no longer explained by HSI score.		
10%	Null model better fit to data. *T. cristatus* occupancy no longer explained by HSI score.		
30%	Null model better fit to data. *T. cristatus* occupancy no longer explained by HSI score.		
Species-specific	3.020 ± 0.791 χ^2_1 = 15.709 ***P* < 0.001**	χ^2_{501} = 506.763 P = 0.420	χ^2_8 = 8.118 P = 0.422 R^2 = 4.99%

Table S3.8: Summary of abiotic determinants of vertebrate species richness as identified using a Poisson GLMM for different metabarcoding sequence thresholds (n = 504 ponds). For categorical variables with more than one level, effect size and standard error are only given for levels reported in the model summary. Test statistic is for LRT used. Significant P-values (<0.05) are in bold.

Threshold	Model overdispersion	Model fit	Model variables	Effect size	Standard error	χ^2	*P*
No threshold	χ^2_{498} = 375.433	χ^2_{8} = -69.777	**Overhang**	-0.002	0.001	10.935	**0.001**
	P = 0.999	P = 1.000	**Rough grass**			8.205	**0.017**
		R^2 = 6.66%	None	0.062	0.002		
			Some	-0.112	0.002		
			Outflow	0.200	0.002	10.988	**0.001**
0.05%	χ^2_{496} = 406.722	χ^2_{8} = -62.768	**Overhang**	-0.002	0.001	6.963	**0.008**
	P = 0.999	P = 1.000	**Outflow**	0.163	0.062	6.735	**0.010**
		R^2 = 6.68%	**Rough grass**			7.374	**0.025**
			None	0.009	0.068		
			Some	-0.145	0.065		
			Scrub/hedge			6.722	**0.035**
			None	-0.079	0.131		
			Some	0.139	0.057		
0.1%	χ^2_{496} = 410.479	χ^2_{8} = -62.194	**Overhang**	-0.002	0.001	8.628	**0.003**
	P = 0.998	P = 1.000	**Outflow**	0.161	0.063	6.443	**0.011**
		R^2 = 6.94%	**Rough grass**			6.538	**0.038**
			None	0.006	0.069		
			Some	-0.140	0.066		
			Scrub/hedge			6.891	**0.032**
			None	-0.091	0.134		
			Some	0.141	0.058		
0.5%	χ^2_{496} = 508.449	χ^2_{8} = -1.413	**Overhang**	-0.002	0.001	9.090	**0.003**
	P = 0.340	P = 1.000	**Outflow**	0.152	0.062	5.946	**0.015**
		R^2 = 6.54%	**Rough grass**			7.430	**0.024**
			None	-0.064	0.076		
			Some	-0.184	0.072		
			Overall terrestrial habitat			6.485	**0.039**
			Moderate	0.193	0.078		
			Poor	0.177	0.087		
1%	χ^2_{501} = 470.396	χ^2_{8} = -35.854	**Overhang**	-0.003	0.001	14.810	**<0.001**
	P = 0.833	P = 1.000					
		R^2 = 3.50%					
5%	χ^2_{499} = 378.448	χ^2_{8} = 39.565	**Overhang**	-0.004	0.001	16.921	**<0.001**
	P = 0.999	P = **<0.001**	**Rough grass**			8.126	**0.017**
		R^2 = 7.66%	None	0.061	0.092		
			Some	-0.185	0.093		
10%	χ^2_{501} = 357.332	χ^2_{8} = -238.540	**Overhang**	-0.007	0.001	26.768	**<0.001**
	P = 0.999	P = 1.000					

		$R^2 = 7.68\%$					
30%	$\chi^2_{497} = 341.011$	$\chi^2_{8} = 10.709$	**Overhang**	-0.011	0.002	25.478	**<0.001**
	$P = 1.000$	$P = 0.219$	**Waterfowl**			7.493	**0.024**
		$R^2 = 12.65\%$	Major	-1.169	0.513		
			Minor	-0.122	0.149		
			Woodland			6.289	**0.043**
			None	-0.448	0.185		
			Some	-0.146	0.179		
Species-specific	$\chi^2_{494} = 431.959$	$\chi^2_{8} = -42.708$	**Outflow**	0.214	0.063	11.220	**0.001**
	$P = 0.979$	$P = 1.000$	**Rough grass**			16.715	**<0.001**
		$R^2 = 8.94\%$	None	-0.140	0.0795		
			Some	-0.297	0.074		
			Overall terrestrial habitat			8.244	**0.016**
			Poor	0.115	0.089		
			Moderate	0.216	0.078		
			Overhang	-0.003	0.0008	9.575	**0.002**
			Macrophyte cover	-0.002	0.001	4.117	**0.043**
			Pond density	0.006	0.003	4.564	**0.033**

Table S3.9: Summary of relationship between HSI score and vertebrate species richness as determined using a Poisson GLMM for different metabarcoding sequence thresholds (n = 504 ponds). Test statistic is for LRT used. Significant P-values (<0.05) are in bold.

Threshold	GLMM results	Overdispersion	Model fit
None	0.474 ± 0.192 χ^2_1 = 6.102 ***P* = 0.014**	χ^2_{501} = 355.432 P = 0.999	χ^2_8 = -109.49 P = 1.000 R^2 = 1.29%
0.05%	0.496 ± 0.002 χ^2_1 = 6.244 ***P* = 0.013**	χ^2_{501} = 380.354 P = 0.999	χ^2_8 = -125.06 P = 1.000 R^2 = 1.35%
0.1%	0.504 ± 0.002 χ^2_1 = 6.251 ***P* = 0.012**	χ^2_{501} = 382.557 P = 0.999	χ^2_8 = -130.31 P = 1.000 R^2 = 1.36%
0.5%	0.472 ± 0.198 χ^2_1 = 5.732 ***P* = 0.017**	χ^2_{501} = 447.442 P = 0.769	χ^2_8 = -42.281 P = 1.000 R^2 = 1.32%
1%	0.561 ± 0.210 χ^2_1 = 7.267 ***P* = 0.007**	χ^2_{501} = 473.185 P = 0.809	χ^2_8 = -5.908 P = 1.000 R^2 = 1.73%
5%	0.683 ± 0.277 χ^2_1 = 6.193 ***P* = 0.013**	χ^2_{501} = 389.934 P = 0.999	χ^2_8 = -47.496 P = 1.000 R^2 = 1.64%
10%	0.897 ± 0.336 χ^2_1 = 7.292 ***P* = 0.007**	χ^2_{501} = 370.163 P = 0.999	χ^2_8 = 126.330 ***P* < 0.001** R^2 = 2.13%
30%	1.189 ± 0.546 χ^2_1 = 4.894 ***P* = 0.027**	χ^2_{501} = 350.580 P = 0.999	χ^2_8 = 10.472 P = 0.233 R^2 = 2.03%
Species-specific	0.459 ± 0.002 χ^2_1 = 4.894 ***P* = 0.025**	χ^2_{501} = 389.744 P = 0.999	χ^2_8 = -145.120 P = 1.000 R^2 = 1.10%

Table S3.10: Summary of species detected by eDNA metabarcoding of freshwater ponds ($N = 532$).

Common name	Binomial name	No. ponds detected
European eel	*Anguilla anguilla*	15
Common barbel	*Barbus barbus*	2
Crucian carp	*Carassius carassius*	2
Common carp	*Cyprinus carpio*	41
Common minnow	*Phoxinus phoxinus*	13
Common roach	*Rutilus rutilus*	72
European chub	*Squalius cephalus*	21
Stone loach	*Barbatula barbatula*	15
Northern pike	*Esox lucius*	17
European bullhead	*Cottus gobio*	14
Three-spined stickleback	*Gasterosteus aculeatus*	56
Ninespine stickleback	*Pungitius pungitius*	15
Ruffe	*Gymnocephalus cernua*	1
Rainbow trout	*Oncorhynchus mykiss*	3
Common toad	*Bufo bufo*	42
Marsh frog	*Pelophylax ridibundus*	1
Common frog	*Rana temporaria*	120
Palmate newt	*Lissotrition helveticus*	5
Smooth newt	*Lissotriton vulgaris*	152
Great crested newt	*Triturus cristatus*	149
Dabbling ducks	*Anas* spp.	7
Eurasian oystercatcher	*Haematopus ostralegus*	1
Common buzzard	*Buteo buteo*	4
Common pheasant	*Phasianus colchicus*	25
Domesticated turkey	*Meleagris gallopavo*	11
Helmeted guineafowl	*Numida meleagris*	1
Eurasian coot	*Fulica atra*	48
Common moorhen	*Gallinula chloropus*	215

Eurasian jay	*Garrulus glandarius*	7
European goldfinch	*Carduelis carduelis*	1
Dunnock	*Prunella modularis*	4
Eurasian nuthatch	*Sitta europaea*	1
Common starling	*Sturnus vulgaris*	4
Melodius warbler	*Hippolais polyglotta*	2
Grey heron	*Ardea cinerea*	1
Great spotted woodpecker	*Dendrocopus major*	1
Green woodpecker	*Picus viridis*	2
Tawny owl	*Strix aluco*	1
Dog	*Canis lupus familiaris*	65
Red fox	*Vulpes vulpes*	9
Eurasian otter	*Lutra lutra*	1
European badger	*Meles meles*	7
European polecat	*Mustela putorius*	1
Common pipistrelle	*Pipistrellus pipistrellus*	1
Eurasian water shrew	*Neomys fodiens*	9
Common shrew	*Sorex araneus*	1
European hare	*Lepus europaeus*	1
European rabbit	*Oryctolagus cuniculus*	24
Horse	*Equus caballus*	3
European water vole	*Arvicola amphibius*	16
Bank vole	*Myodes glareolus*	9
House mouse	*Mus musculus*	16
Brown rat	*Rattus norvegicus*	39
Grey squirrel	*Sciurus carolinensis*	57
Cow	*Bos taurus*	179
Sheep	*Ovis aries*	42
Red deer	*Cervus elaphus*	2
Reeve's muntjac	*Muntiacus reevesi*	3
Pig	*Sus scrofa domesticus*	140
Cat	*Felis catus*	16

Appendix 3.4: Supplementary figures

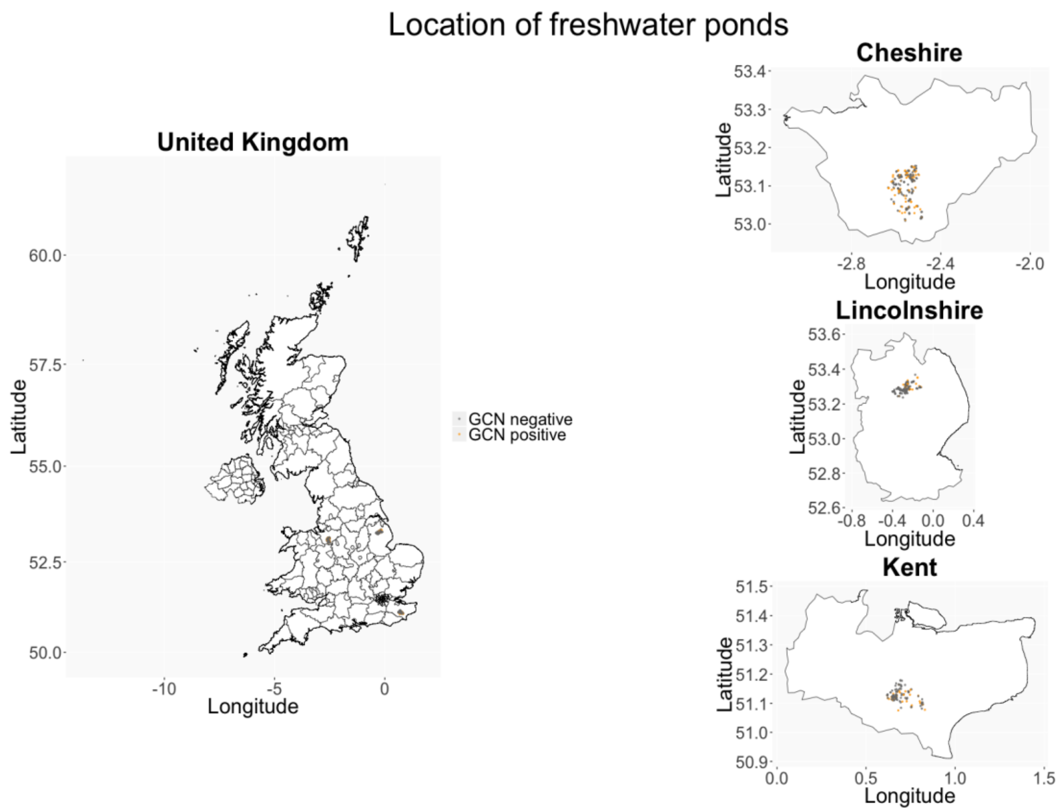


Figure S3.1: Location of ponds (*n* = 504) sampled for eDNA as part of Natural England's Great Crested Newt Evidence Enhancement Programme. Ponds that were negative or positive for *T. cristatus* (GCN) by targeted quantitative PCR are indicated by grey and orange points respectively.

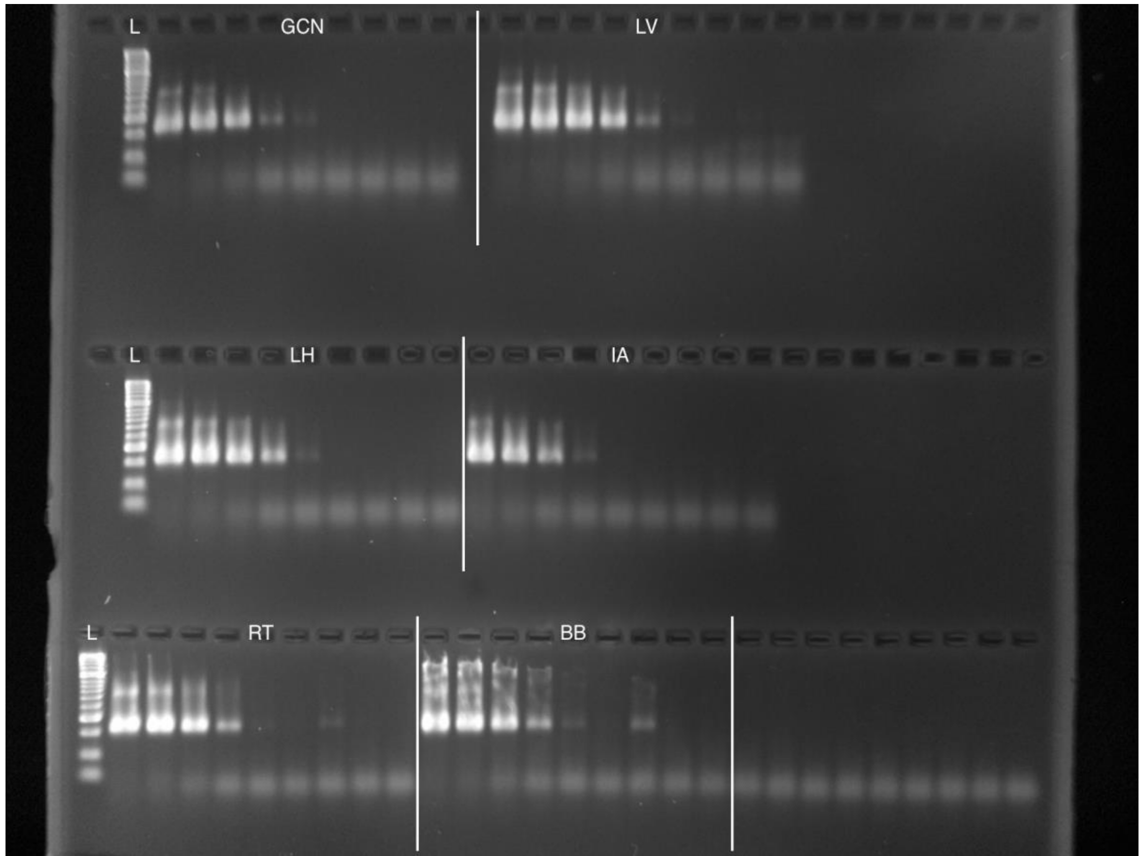


Figure S3.2: Gel image showing results of *in vitro* primer validation. All tissue DNA used for dilution series was standardised to a starting concentration of 5 ng/µl. The Limit of Detection was variable for each species: *Triturus cristatus* (GCN), *Lissotriton helveticus* (LH), *Rana temporaria* (RT) and *Bufo bufo* (BB) were not amplified below 5 x 10^{-4} ng/µl, whereas *Icthyosaura alpestris* (IA) was was not amplified below 5 x 10^{-3} ng/µl and *Lissotriton vulgaris* (LV) below 5 x 10^{-5} ng/µl.

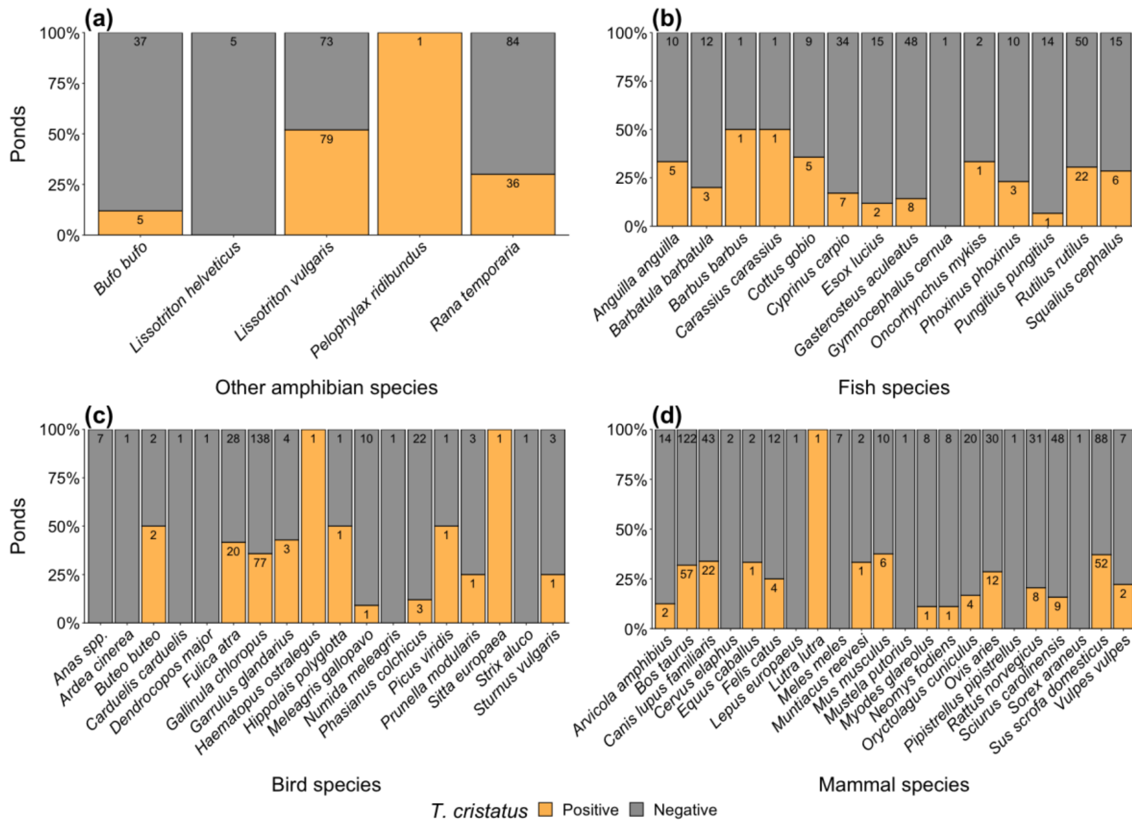


Figure S3.3: Occurrence of *T. cristatus* in relation to species from different vertebrate groups (N = 532 ponds): **(a)** other amphibians, **(b)** fish, **(c)** birds, and **(d)** mammals. Numbers on each bar are the number of ponds with and without *T. cristatus* in which a species was detected.

Appendix 3.5: Supplementary references

Bjørnstad, O.N. (2017) ncf: Spatial Covariance Functions. R package version 1.1-7.

Edgar, P. & Bird, D.R. (2006) *Action plan for the conservation of the crested newt Triturus cristatus species complex in Europe*. Council of the European Union, Strassbourg, Germany.

Fox, J. & Weisberg, S. (2011) *An R Companion to Applied Regression, Second Edition*. Sage, Thousand Oaks, CA.

Griffith, D., Veech, J. & Marsh, C. (2016) cooccur: Probabilistic Species Co-Occurrence Analysis in R. *Journal of Statistical Software*, **69**, 1–17.

Haubrock, P.J. & Altrichter, J. (2016) Northern crested newt (*Triturus cristatus*) migration in a nature reserve: multiple incidents of breeding season displacements exceeding 1km. *The Herpetological Bulletin*, **138**, 31–33.

Therneau, T., Atkinson, B. & Ripley, B. (2014) rpart: Recursive Partitioning and Regression Trees. R package version 4.1-13.

Veech, J.A. (2013) A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, **22**, 252–260.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., & Smith, G.M. (2009) *Mixed effects models and extensions in ecology with R*. Springer, New York, USA.

Appendix 4

Appendix 4.1: Supplementary methods

Assay optimisation

Conventional PCR

Primers were validated *in vitro* using PCR and tissue DNA (standardised to 1 ng/µL) from fin clips of crucian carp (*Carassius carassius*) and four closely related non-target species: goldfish (*Carassius auratus*), common carp (*Cyprinus carpio*), tench (*Tinca tinca*), and sunbleak (*Leucaspius delineatus*). An annealing temperature gradient was performed with target and non-target DNA to test assay specificity. The cycling conditions detailed below were used but with annealing temperatures of 48 °C, 50 °C, 52 °C, 54 °C, 56 °C, and 58 °C (Figure S4.1). The gradient PCR revealed all subsequent PCRs should be performed at an annealing temperature of 60 °C. Primers were then tested on eDNA samples from ponds recently stocked with *C. carassius* to confirm potential for eDNA amplification (Figure S4.2). All PCR reactions were performed in 20 µL volumes containing: 10 µL of MyTaq™ Red Mix (Bioline®, UK), 1 µL of forward and reverse primer (0.04 µM) (Integrated DNA Technologies, Belgium), 6 µL of molecular grade sterile water (Fisher Scientific UK Ltd, UK) and 2 µL of DNA template. PCRs were conducted on an Applied Biosystems® Veriti Thermal Cycler with the following profile: 95 °C for 3 min, 35 cycles of 95 °C for 30 sec, 60 °C for 60 sec and 72 °C for 90 sec, followed by a final elongation step at 72 °C for 10 min. Molecular grade water (Fisher Scientific UK Ltd, UK) was used as the no template control (NTC) in all tests. PCR products were stored at 4 °C until fragment size was verified by visualising 2 µL of selected PCR product on 2% agarose gels (80 mL 0.5x TBE buffer, 1.6 g agarose powder). Gels were then stained with ethidium bromide or GelRed™ (VWR International, UK), and imaged using Image Lab™ Software (Bio-Rad Laboratories Ltd, UK). A PCR product was deemed positive where there was an amplification band on the gel that was of the expected size (118 bp).

Primer and probe concentration

A matrix of primer concentrations were tested on a single qPCR reaction plate, ranging from 50 nM to 900 nM in final concentration. The probe concentration remained constant at a final concentration of 125 nM. The primer concentrations resulting in the lowest cycle threshold value (C_q) and high endpoint fluorescence relative to the most concentrated level tested were considered optimal (Bustin *et al*., 2009; Wilcox *et al*., 2015; Dysthe *et al*., 2018). These optimised primer concentrations were then used to determine the optimal probe concentration. Probe reaction concentrations of 50nM, 125nM, 150nM, 200nM, 250nM, 300nM, 400nM, and 500nM were tested on 10 copies/µL of gBlocks® gene fragment to assess change in probe sensitivity. The lowest concentration of probe which allows the most sensitive detection (lowest C_q value with consistent amplification) is optimal. The optimal concentrations were adopted for subsequent qPCR analyses performed to determine assay specificity and sensitivity, and quantify eDNA samples.

Standard curve preparation and cycle number

We encountered problems with qPCR amplification efficiency and pipetting accuracy (R^2). Tests continually showed poor amplification efficiency until we processed samples using optical 96-well plates and seals (Applied Biosystems™, UK) instead of optical strip tubes and caps (Applied Biosystems™, UK). We re-optimised the assay on plates and began to process eDNA samples, but then experienced problems with our gBlocks® Gene Fragment (Integrated DNA Technologies, Belgium) standard curve. The standard curve did not amplify consistently and was not reproducible between plates. Upon running a four-way comparison of standard curve preparation and primer concentrations, we discovered that our assay was most efficient using the primer concentrations initially identified, but preparing standard curve dilutions fresh rather than freeze-thawing aliquots. For all subsequent plates (including assay validation and analysis of eDNA samples), we aliquoted 2 µL of gBlocks® (10^7 copies/µL) into the first tube on a PCR tube strip, and froze this at -20 °C. When setting up a qPCR plate, we removed a single PCR tube strip to thaw and added 18 µL of IDTE buffer (pH 8.0) (Integrated DNA Technologies, Belgium), from which subsequent 10-fold dilutions were made. Standards were not reused for qPCRs. Initial optimisation tests in qPCR indicated that 60 cycles were required for the amplification curve of our lowest standard (10 copies/µL) to plateau.

However, with changes to preparation of standards, the amplification curve of our lowest standard consistently plateaued within 45 reaction cycles. Although we continued to use 60 cycles for consistency with earlier tests, 45 cycles could be used for subsequent applications. This cycle number would allow the amplification curve of the 1 copy/μL standard (mean C_q value = 40.07) to grow and plateau.

Conventional PCR Vs qPCR

We performed a small-scale comparison of qPCR and conventional PCR for *C. carassius* detection, where qPCR was hypothesised to possess greater detection sensitivity than PCR. All five eDNA samples from two ponds (RAIL and MYST) were analysed by PCR and qPCR using the same number of technical replicates and a standard curve for quantification. PCR conditions were as described in 'Conventional PCR', whereas qPCR conditions are detailed in Chapter 4.

Data analysis

Variation in DNA copy number between eDNA samples

We examined variation in DNA copy number amongst samples for each pond using a hierarchical, Poisson Generalized Linear Mixed Model (GLMM) within the R package glmmTMB v0.2.0 (Brooks *et al.*, 2017). Prior to modelling, all variables were assessed for collinearity using Spearman's correlation coefficient and Variance Inflation Factors (VIFs) calculated using the R package car v2.1-6 (Fox & Weisberg, 2011). Variables were considered collinear and removed if r >0.3 and VIF >3 (Zuur *et al.*, 2009). Number of filters used was removed on this basis but volume of water filtered was not given marginal r >0.3 with most variables but VIF <3. Replicates nested within each pond were modelled as a random effect, whilst volume of water filtered, Qubit concentration, and presence of sediment, vegetation, and algae in water samples were modelled as fixed effects. Validation checks were performed to ensure all model assumptions were met and absence of overdispersion (Zuur *et al.*, 2009). Model fit was assessed visually and with the Hosmer and Lemeshow Goodness of Fit Test (Hosmer & Lemeshow, 2000) using the R package ResourceSelection v0.3-0 (Lele *et al.*, 2014). Model predictions were obtained using the *predict* function and upper and lower 95% CIs were calculated from the standard

error of the predictions. All values were bound in a new data frame and model results plotted for evaluation using the R package ggplot2 v2.2.1 (Wickham, 2016).

Appendix 4.2: Supplementary results and discussion

Variation in DNA copy number between eDNA samples

C. carassius detection and DNA copy number were highly variable across the five biological replicates collected from some ponds (Figure S4.6). A GLMM (model fit: $\chi^2_8 = 0.443$, $P = 0.999$) identified Qubit concentration (0.031 ± 0.012, $\chi^2_1 = 6.614$, $P = 0.010$), presence of duckweed in water samples (3.106 ± 1.032, $\chi^2_1 = 7.263$, $P = 0.007$), and presence of sediment in water samples (-2.472 ± 1.164, $\chi^2_1 = 4.221$, $P = 0.040$) as significant predictors of DNA copy number, where DNA copy number increased as Qubit concentration increased and where duckweed (*Lemna* spp.) was present in samples, but decreased where sediment was present in samples (Figure S4.7).

PCR vs. qPCR

C. carassius eDNA was amplified by PCR in all samples that amplified using qPCR (Table S4.4). PCR also provided semi-quantitative estimates of eDNA concentration when PCR products for eDNA samples were run on gels alongside qPCR standards (Figure S4.8). Our study is not the first to compare eDNA detection using different means of DNA amplification (Nathan *et al.*, 2014; Farrington *et al.*, 2015; Piggott, 2016; De Ventura *et al.*, 2017). Like Nathan *et al.* (2014), we found PCR had comparable sensitivity to qPCR and band strength of PCR products may indicate eDNA concentration; however, we also translated band strength to approximate DNA copy number. PCR may require more replication to achieve set detection probabilities (Piggott, 2016), but lower sensitivity could make this approach more robust to false positives from cross-contamination than qPCR (De Ventura *et al.*, 2017). Large-scale comparisons of PCR and qPCR across study systems and species are needed to truly assess performance of each approach. Nonetheless, our findings support PCR as a cost-efficient, semi-quantitative alternative to qPCR for conservation programmes wishing to utilise eDNA (Nathan *et al.*, 2014; De Ventura *et al.*, 2017).

Appendix 4.3: Supplementary tables

Table S4.1: Summary of eDNA analysis for each sample collected from ponds in Norfolk, eastern England, including volume of water filtered, number of filters used, qPCR result, copy number when originally analysed, and copy number when spiked with synthetic target DNA (1000 copies/µL) for inhibition testing. Copy numbers of partially inhibited samples are highlighted in red.

Pond	*C. carassius* (Y/N)	Sample	Volume filtered (L)	No. of filters used	qPCR amplific-ation	Non-spiked DNA copy number (copies/µL)	Spiked DNA copy number (copies/µL)
SABA	N	1	1	1	N	0	1260
		2	1	1	N	0	1271
		3	1	1	N	0	1362
		4	1	1	N	0	1406
		5	1	1	N	0	1238
WRONG	N	1	1	2	N	0	1247
		2	1	2	N	0	1187
		3	1	2	N	0	1277
		4	1	2	N	0	1409
		5	1	2	N	0	1281
WADD10	N	1	1	1	N	0	1092
		2	1	1	N	0	1891
		3	1	1	N	0	1207
		4	1	1	N	0	1233
		5	1	1	N	0	1264
WADD11	N	1	0.5	2	N	0	1308
		2	0.5	2	N	0	1657
		3	0.5	2	N	0	1386
		4	0.5	2	N	0	1268
		5	0.5	2	N	0	1287

WADD17	N	1	1	1	N	0	1015
		2	1	2	N	0	1314
		3	1	1	N	0	1262
		4	1	2	N	0	1216
		5	1	2	N	0	1324
WOOD	N	1	1	1	N	0	1473
		2	1	2	N	0	1204
		3	1	2	N	0	1177
		4	1	2	N	0	1353
		5	1	2	N	0	1360
PYES2	N	1	1	1	N	0	1351
		2	1	1	N	0	1299
		3	1	1	N	0	1347
		4	1	1	N	0	1571
		5	1	2	N	0	887
VALE	N	1	0.5	2	N	0	1665
		2	0.5	2	N	0	1548
		3	0.5	2	N	0	1632
		4	0.5	2	N	0	1647
		5	0.5	2	N	0	1671
LDUN2	N	1	1	1	N	0	1411
		2	1	1	N	0	1526
		3	1	1	N	0	1480
		4	1	1	N	0	1538
		5	1	1	N	0	1623
LDUN3	N	1	1	1	N	0	1441
		2	1	1	N	0	1645
		3	1	2	N	0	1657
		4	1	2	N	0	1611
		5	1	2	N	0	2056

SKEY1	Y	1	1	1	N	0	1345
		2	1	1	N	0	1209
		3	1	1	Y	17	1304
		4	1	1	Y	95	1526
		5	1	1	Y	21	1342
OTOM	Y	1	1	1	Y	91	1701
		2	1	1	Y	57	1428
		3	1	1	Y	81	1585
		4	1	1	Y	143	1649
		5	1	1	Y	127	1546
CHIP	Y	1	1	1	N	0	1382
		2	1	1	N	0	1724
		3	1	1	N	0	1482
		4	1	1	N	0	1566
		5	1	1	N	0	1574
GUES1	Y	1	1	1	Y	116	1529
		2	1	1	Y	91	1466
		3	1	1	N	0	1303
		4	1	1	Y	62	1680
		5	1	1	Y	158	738
WADD3	Y	1	1	1	Y	128	1361
		2	1	1	N	0	1528
		3	1	1	Y	179	1527
		4	1	1	Y	407	1765
		5	1	1	Y	341	1664
POHI	Y	1	1	1	Y	5	1330
		2	1	1	Y	2	1413
		3	1	1	Y	1	1110
		4	1	1	Y	4	1302
		5	1	1	N	0	774

POFA4	Y	1	1	1	Y	8	1444
		2	1	1	N	0	1401
		3	1	1	Y	49	1562
		4	1	1	Y	7	1392
		5	1	1	Y	79	1421
RAIL	Y	1	1	1	Y	39	404
		2	1	1	N	0	1582
		3	1	1	Y	153	1778
		4	1	1	Y	230	1918
		5	1	1	Y	43	1735
MYST	Y	1	1	1	Y	5	1701
		2	1	1	Y	5	1428
		3	1	1	Y	9	1585
		4	1	1	Y	7	1649
		5	1	1	Y	6	1546
CAKE	Y	1	0.25	2	Y	2	1584
		2	0.25	2	Y	2	1426
		3	0.25	2	N	0	1253
		4	0.25	2	Y	2	1375
		5	0.25	2	Y	2	1211

Table S4.2: Corresponding species, accession number, and geographic location for sequences that were downloaded from the NCBI nucleotide database to construct an alignment of consensus sequences for assay design.

Species	Accessions	Geographic location
Carassius carassius	AY714387.1	China
	DQ399917.1	Germany
	DQ399918.1	Germany
	DQ399919.1	Germany
	DQ399938.1	Czech Republic
	FJ167428.1	Europe
	GU991399.1	Czech Republic
	GU991400.1	UK
	HQ689908.1	Russia
	HQ689909.1	Russia
	JN412533.1	Austria
	JN412534.1	Austria
	JN412535.1	Austria
	JN412536.1	Austria
	JN412537.1	Germany
	JN412538.1	Germany
	JN412539.1	UK
	JN412540.1	Germany
	JN412541.1	Germany
	JN412542.1	Germany
	JN412543.1	Germany
	JN412544.1	Sweden
	JN412545.1	Sweden
	JN412546.1	Sweden
	JN412547.1	Sweden
	JN412548.1	Sweden
	JN412549.1	Czech Republic
	JN412550.1	Czech Republic
	JQ763597.1	Czech Republic
	KC238569.1	Czech Republic
	KR131834.1	Czech Republic
	KR131835.1	Czech Republic
	KR131836.1	Czech Republic
	KR131837.1	Czech Republic
	KR131838.1	Czech Republic
	KR131839.1	Czech Republic
	KR131840.1	Czech Republic

KR131841.1	Czech Republic
KR131842.1	Czech Republic
KR131843.1	Finland
KR131844.1	Germany
KR131845.1	Germany
KT630314.1	Finland
KT630315.1	Finland
KT630316.1	Finland
KT630317.1	Estonia
KT630318.1	Germany
KT630319.1	Estonia
KT630320.1	Germany
KT630321.1	Sweden
KT630322.1	Sweden
KT630323.1	Sweden
KT630324.1	Belgium
KT630325.1	Estonia
KT630326.1	Germany
KT630327.1	Germany
KT630328.1	Germany
KT630329.1	Norway
KT630330.1	Norway
KT630331.1	Sweden
KT630332.1	Sweden
KT630333.1	Sweden
KT630334.1	Russia
KT630335.1	Russia
KT630336.1	Russia
KT630337.1	Finland
KT630338.1	Finland
KT630339.1	Finland
KT630340.1	Finland
KT630341.1	Finland
KT630342.1	Finland
KT630343.1	Poland
KT630344.1	Poland
KT630345.1	Poland
KT630346.1	Russia
KT630347.1	Russia
KT630348.1	Russia
KT630349.1	Sweden
KT630350.1	Sweden
KT630351.1	Sweden

KT630352.1	Poland
KT630353.1	Poland
KT630354.1	Poland
KT630355.1	Sweden
KT630356.1	Sweden
KT630357.1	Sweden
KT630358.1	Russia
KT630359.1	Russia
KT630360.1	Russia
KT630361.1	Russia
KT630362.1	Russia
KT630363.1	Belarus
KT630364.1	Russia
KT630365.1	Sweden
KT630366.1	Sweden
KT630367.1	Sweden
KT630368.1	Sweden
KT630369.1	Sweden
KT630370.1	Sweden
KT630371.1	Sweden
KT630372.1	Sweden
KT630373.1	Sweden
KT630374.1	UK
KT630375.1	UK
KT630376.1	UK
KT630377.1	UK
KT630378.1	UK
KT630379.1	UK
KT630380.1	UK
KT630381.1	Czech Republic
KT630382.1	Czech Republic
KT630383.1	Czech Republic
KT630384.1	Germany
KT630385.1	Germany
KT630386.1	Germany
KT630387.1	Germany
KT630388.1	Finland
KT630389.1	Finland
KT630390.1	Finland
KT630391.1	Hungary
KT630392.1	Germany
KT630393.1	Germany
KT630394.1	Germany

	NC_006291.1	China
Carassius auratus	AB368677.1	Ryukyu Archipelago
	AB368678.1	Ryukyu Archipelago
	AB368679.1	Ryukyu Archipelago
	AB368680.1	Ryukyu Archipelago
	AB368681.1	Ryukyu Archipelago
	AB368682.1	Ryukyu Archipelago
	AB368683.1	Ryukyu Archipelago
	AB368684.1	Ryukyu Archipelago
	AB368685.1	Ryukyu Archipelago
	AB368686.1	Ryukyu Archipelago
	AB368687.1	Ryukyu Archipelago
	AB368688.1	Ryukyu Archipelago
	AB368689.1	Ryukyu Archipelago
	AB368690.1	Ryukyu Archipelago
	AB368691.1	Ryukyu Archipelago
	AB368692.1	Ryukyu Archipelago
	AB368693.1	Ryukyu Archipelago
	AB368694.1	Ryukyu Archipelago
	AB368695.1	Ryukyu Archipelago
	AB368696.1	Ryukyu Archipelago
	AB368697.1	Ryukyu Archipelago
	AB368698.1	Ryukyu Archipelago
	AB368699.1	Ryukyu Archipelago
	AB368700.1	Ryukyu Archipelago
	AB368701.1	Ryukyu Archipelago
	AB368702.1	Ryukyu Archipelago
	AB368703.1	Ryukyu Archipelago
	AB368704.1	Ryukyu Archipelago
	AB368705.1	Ryukyu Archipelago
	AB368706.1	Ryukyu Archipelago
	AB368707.1	Ryukyu Archipelago
	AB368708.1	Ryukyu Archipelago
	AB368709.1	Ryukyu Archipelago
	AB368710.1	Ryukyu Archipelago
	AB379915.1	Japan
	AB379916.1	Japan
	AB379917.1	Japan
	AB379918.1	Japan
	AB379919.1	Japan
	AB379920.1	Japan
	AB379921.1	China
	AF045966.1	Europe

AP011239.1	Japan
AP017363.1	Japan
AP017364.1	Japan
AP017365.1	Japan
DQ399920.1	Japan
DQ399921.1	Japan
DQ399922.1	Japan
DQ399923.1	Czech Republic
DQ399924.1	Czech Republic
DQ399925.1	Czech Republic
DQ399930.1	Czech Republic
DQ399932.1	Czech Republic
DQ868897.1	Czech Republic
DQ868898.1	Czech Republic
DQ868906.1	Portugal
DQ868907.1	Portugal
DQ868908.1	Portugal
DQ868909.1	Spain
DQ868913.1	Spain
DQ868914.1	Spain
DQ868915.1	Spain
DQ868921.1	Portugal
DQ868922.1	Portugal
DQ868923.1	Portugal
DQ868927.1	Czech Republic
DQ868928.1	Czech Republic
EU364877.1	China
EU663597.1	China
EU663598.1	China
EU663599.1	China
FJ169952.1	Europe
FJ169953.1	Europe
FJ169954.1	Europe
GU135503.1	China
GU135504.1	China
GU942707.1	Germany
GU942708.1	Germany
GU942709.1	Bosnia and Herzegovina
GU991382.1	Portugal
GU991383.1	Portugal
GU991384.1	Portugal
GU991385.1	Czech Republic
GU991386.1	Albania

GU991387.1	Albania
GU991388.1	Montenegro
GU991389.1	Montenegro
GU991390.1	Albania
GU991391.1	Kazakhstan
GU991392.1	China
GU991393.1	Greece
GU991394.1	Greece
GU991395.1	Greece
GU991396.1	Greece
GU991397.1	Kazakhstan
GU991398.1	Korea
HM000036.1	Italy
HM008691.1	Italy
HQ443698.1	Taiwan
HQ689793.1	China
HQ689794.1	China
HQ689795.1	China
HQ689796.1	China
HQ689797.1	China
HQ689798.1	China
HQ689799.1	China
HQ689800.1	China
HQ689801.1	China
HQ689802.1	China
HQ689803.1	China
HQ689804.1	China
HQ689805.1	China
HQ689806.1	China
HQ689807.1	China
HQ689808.1	China
HQ689809.1	China
HQ689810.1	China
HQ689811.1	China
HQ689812.1	China
HQ689813.1	China
HQ689814.1	China
HQ689815.1	China
HQ689816.1	China
HQ689817.1	Japan
HQ689818.1	China
HQ689819.1	China
HQ689820.1	China

HQ689821.1	China
HQ689822.1	China
HQ689823.1	China
HQ689824.1	China
HQ689825.1	China
HQ689826.1	China
HQ689827.1	China
HQ689828.1	China
HQ689829.1	China
HQ689830.1	China
HQ689831.1	China
HQ689832.1	China
HQ689833.1	China
HQ689834.1	China
HQ689835.1	China
HQ689836.1	China
HQ689837.1	China
HQ689838.1	China
HQ689839.1	China
HQ689840.1	China
HQ689841.1	China
HQ689842.1	China
HQ689843.1	China
HQ689844.1	China
HQ689845.1	China
HQ689846.1	China
HQ689847.1	China
HQ689848.1	China
HQ689849.1	China
HQ689850.1	China
HQ689851.1	China
HQ689852.1	China
HQ689853.1	China
HQ689854.1	China
HQ689855.1	China
HQ689856.1	China
HQ689857.1	China
HQ689858.1	China
HQ689859.1	China
HQ689860.1	China
HQ689861.1	China
HQ689862.1	China
HQ689863.1	China

HQ689864.1	China
HQ689865.1	China
HQ689866.1	China
HQ689867.1	China
HQ689868.1	China
HQ689869.1	China
HQ689870.1	China
HQ689871.1	China
HQ689872.1	China
HQ689873.1	China
HQ689874.1	China
HQ689875.1	Vietnam
HQ689876.1	Vietnam
HQ689877.1	Vietnam
HQ689878.1	China
HQ689879.1	China
HQ689880.1	China
HQ689881.1	China
HQ689882.1	China
HQ689883.1	China
HQ689884.1	China
HQ689885.1	China
HQ689886.1	China
HQ689887.1	China
HQ689888.1	China
HQ689889.1	China
HQ689890.1	China
HQ689910.1	Japan
HQ689911.1	Japan
HQ689912.1	Japan
HQ875340.1	China
JF694778.1	Russia
JF694779.1	Russia
JF694780.1	Russia
JN105355.1	China
JN412507.1	Austria
JN412508.1	China
JN412509.1	China
JN412510.1	China
JN412511.1	China
JN412512.1	China
JN412513.1	Illinois, USA
JN412514.1	Illinois, USA

JN412515.1	Illinois, USA
JN412516.1	Illinois, USA
JN412517.1	Illinois, USA
JN412518.1	Germany
JN412519.1	China
JN412520.1	China
JN412521.1	Germany
JN412522.1	Germany
JN412523.1	China
JN412524.1	China
JN412525.1	Montenegro
JN412526.1	Montenegro
JN412527.1	Japan
JN412528.1	Ukraine
JN412529.1	Czech Republic
JN412530.1	Czech Republic
JN412531.1	Ukraine
JN412532.1	Ukraine
KF147851.1	China
KF731743.1	Czech Republic
KF731744.1	Czech Republic
KF731745.1	Czech Republic
KJ476998.1	China
KJ874428.1	China
KJ874429.1	China
KM015475.1	Russia
KM015476.1	Russia
KM261774.1	North Korea
KM657141.1	Turkey
KM657142.1	Turkey
KM657143.1	Turkey
KM659025.1	China
LC097470.1	Japan
LC097471.1	Japan
LC097472.1	Japan
LC097473.1	Japan
LC097474.1	Japan
LC097475.1	Japan
LC097476.1	Japan
LC097477.1	Japan
LC097478.1	Japan
LC097914.1	Japan
LC097915.1	Japan

	LC097916.1	Japan
	LC097917.1	Japan
	LC097918.1	Japan
	LC097919.1	Japan
	LC097920.1	Japan
	LC097921.1	Japan
	LC097922.1	Japan
	LC097923.1	Japan
	LC097924.1	Japan
	LC097925.1	Japan
	LC097926.1	Japan
	LC097927.1	Japan
	NC_002079.1	Japan
	NC_015142.1	China
Cyprinus carpio	AB158804.1	Japan
	AB158805.1	Japan
	AB158806.1	Japan
	AB158807.1	Japan
	AP009047.1	Japan
	AY347276.1	China
	AY347277.1	China
	AY347278.1	China
	AY347279.1	China
	AY347280.1	China
	AY347281.1	China
	AY347282.1	China
	AY347283.1	Russia
	AY347284.1	China
	AY347285.1	Japan
	AY347286.1	China
	AY347287.1	Russia
	AY347288.1	China
	AY347289.1	Japan
	AY347290.1	China
	AY347291.1	China
	AY347292.1	Germany
	AY347293.1	Russia
	AY347294.1	Russia
	AY347295.1	Russia
	DQ532100.1	Vietnam
	DQ532101.1	Vietnam
	DQ532102.1	Vietnam
	DQ532103.1	Vietnam

DQ532104.1	Vietnam
DQ532105.1	Vietnam
DQ532106.1	Vietnam
DQ532107.1	Vietnam
DQ532108.1	Indonesia
DQ532109.1	Indonesia
DQ532110.1	China
DQ532111.1	India
DQ532112.1	India
DQ532113.1	Czech Republic
DQ532114.1	Hungary
DQ532115.1	Israel
DQ868871.1	Greece
DQ868872.1	Greece
DQ868873.1	Greece
DQ868874.1	Greece
DQ868875.1	Greece
EU676848.1	Oregon, USA
EU689059.1	Greece
EU689060.1	Greece
EU689061.1	Greece
EU689062.1	Greece
EU689063.1	Greece
EU689064.1	Greece
EU689065.1	Greece
EU689066.1	Greece
EU689067.1	Greece
EU689068.1	Greece
EU689069.1	Greece
EU689070.1	Greece
EU689071.1	Greece
EU689072.1	Greece
FJ478020.1	Russia
FJ478021.1	RussiaChina
HM008692.1	Thailand
HQ443697.1	Taiwan
JN105352.1	China
JN105353.1	China
JN105354.1	China
KF574485.1	India
KF574486.1	India
KF574487.1	India
KF574488.1	India

	KF574489.1	India
	KF574490.1	India
	KF856964.1	China
	KF856965.1	China
	KJ511882.1	Japan
	KJ511883.1	Hungary
	KP013086.1	Indiana, USA
	KP993136.1	China
	KP993137.1	China
	KP993138.1	Russia/China
	KP993139.1	Germany
	KU050703.1	China
	KX710076.1	China
	MG570426.1	Pennsylvania, USA
	MG570427.1	Pennsylvania, USA
	MG570435.1	South Carolina, USA
	NC_018036.1	China
	NC_018037.1	China
Abramis brama	KX588534.1	Czech Republic
	KX588535.1	Czech Republic
	KX588536.1	Czech Republic
	KX588537.1	Czech Republic
	KX588538.1	Czech Republic
	KX588539.1	Czech Republic
	KX588540.1	Czech Republic
	KX588541.1	Czech Republic
	KX588542.1	Czech Republic
	KX588543.1	Czech Republic
	KX588544.1	Czech Republic
	Y10441.1	France
Alburnus alburnus	AF090743.1	Republic of Macedonia/Greece
	AF090744.1	Greece
	AF090745.1	Greece
	AY026393.1	Armenia
	DQ350253.1	Portugal
	DQ350254.1	Croatia
	HM560060.1	Czech Republic
	HM560061.1	Czech Republic
	HM560062.1	Czech Republic
	JQ436541.1	Spain
	Y10443.1	France
Barbatula barbatula	DQ025767.1	Germany

	DQ025768.1	Spain
	DQ025769.1	UK
	DQ025770.1	UK
	DQ025771.1	Russia
	DQ025772.1	Russia
	DQ025773.1	Russia
	DQ025774.1	Russia
	DQ025775.1	Russia
	DQ025776.1	Russia
	DQ025777.1	Russia
	DQ025778.1	Russia
	DQ025779.1	UK
	DQ025780.1	UK
	DQ025781.1	UK
	DQ025782.1	UK
	DQ025783.1	UK
	DQ025784.1	UK
	DQ025785.1	Belgium
	DQ025786.1	Poland
	DQ025787.1	Poland
	DQ025788.1	Poland
	DQ025789.1	Poland
	DQ025790.1	Belgium
	DQ025791.1	Ukraine
	DQ025792.1	Finland
	DQ025793.1	Finland
	DQ025794.1	Finland
	DQ025795.1	Lithuania
	DQ025796.1	Finland
	DQ025797.1	Poland
	DQ025798.1	Russia
	DQ025800.1	Russia
	DQ025801.1	Russia
	DQ025804.1	Russia
	DQ025807.1	Ukraine
	DQ025808.1	Romania
	DQ025809.1	Russia
	DQ025810.1	Ukraine
Barbus barbus	AF090780.1	Greece/Republic of Macedonia
	AF090781.1	Greece
	AY331017.1	Yugoslavia
	AY331018.1	Yugoslavia

	AY331019.1	Bulgaria
	AY331020.1	Bulgaria
	AY331021.1	Czech Republic
	AY331022.1	Moldova
	AY331023.1	Moldova
	AY331024.1	Moldova
	AY004754.1	Greece
	KC465918.1	Italy
	KC465919.1	Italy
	KC465920.1	Italy
	KC465921.1	Italy
	KC465922.1	Italy
	KC465923.1	Italy
	KC465924.1	Italy
	KC465925.1	Italy
	KC465926.1	Italy
	KC465927.1	Italy
Blicca bjoerkna	HM560076.1	Czech Republic
	HM560077.1	Czech Republic
	EF137863.1	Unknown
	Y10442.1	France
Chondrostoma nasus	AF533761.1	Montenegro
	AY026402.1	Austria
	DQ350254.1	Croatia
	DQ447729.1	France
	KF529136.1	Portugal
	Z75109.1	France
Ctenopharyngodon idella	AB900162.1	Japan
	HM237984.1	China
	HM237985.1	China
	HM237986.1	China
	HM237987.1	China
	HM237988.1	China
	HM237989.1	China
	HM237990.1	China
	HM237991.1	China
	HM237992.1	China
	HM237993.1	China
	HM237994.1	China
	HM237995.1	China
	HM237996.1	China
	HM237997.1	China
	HM237998.1	China

HM237999.1	China
HM238000.1	China
HM238001.1	China
HM238002.1	China
HM238003.1	China
HM238004.1	China
HM238005.1	China
HM238006.1	China
HM238007.1	China
HM238008.1	China
HM238009.1	China
HM238010.1	China
HM238011.1	China
HM238012.1	China
HM238013.1	China
HM238014.1	China
HM238015.1	China
HM238016.1	China
HM238017.1	China
HM238018.1	China
HM238019.1	China
HM238020.1	China
HM238021.1	China
HM238022.1	China
HM238023.1	China
HM238024.1	China
HM238025.1	China
HM238026.1	China
HM238027.1	China
HM238028.1	China
HM238029.1	China
HM238030.1	China
HM238031.1	China
HM238032.1	China
HM238033.1	China
HM238034.1	China
HM238035.1	China
HM238036.1	China
HM238037.1	China
HM238038.1	China
HM238039.1	China
HM238040.1	China
HM238041.1	China

	HM238042.1	China
	HM238043.1	China
Gobio gobio	AF045996.1	Spain
	AY426561.1	Spain
	AY426562.1	Spain
	AY426563.1	Spain
	AY426564.1	Spain
	AY426565.1	Spain
	AY426566.1	France
	AY426567.1	France
	AY426568.1	France
	AY426569.1	France
	AY426570.1	Spain
	AY426571.1	France
	AY426572.1	France
	AY426573.1	Spain
	AY426574.1	Spain
	AY426575.1	Spain
	AY426576.1	Spain
	AY426577.1	Spain
	AY426578.1	Spain
	AY426579.1	Spain
	AY426580.1	Spain
	AY426581.1	Spain
	AY426582.1	Spain
	AY426583.1	Spain
	AY426584.1	Spain
	AY426585.1	Spain
	AY426586.1	Spain
	AY426587.1	Spain
	AY426588.1	Spain
	AY426589.1	Spain
	AY426590.1	Spain
	AY426591.1	Spain
	AY426592.1	Czech Republic
Hypophthalmichthys molitrix	AB198974.1	Russia
	AF051866.1	Unknown
Hypophthalmichthys nobilis	JQ346141.1	Laos
	AF051855.1	Unknown
Leucaspius delineatus	HM560097.1	Russia
	Y10447.1	France
Leuciscus cephalus	AF045995.1	Spain
	AF090752.1	Bulgaria/Greece

AF090753.1	Albania/Greece/Republic of Macedonia
AF090754.1	Greece
AF090755.1	Greece
AF095609.1	Russia
AF421792.1	Italy
AF421801.1	Spain
AF421803.1	Spain
AJ252783.1	Europe
AJ252784.1	Europe
AJ252785.1	Europe
AJ252786.1	Europe
AJ252787.1	Europe
AJ252788.1	Europe
AJ252789.1	Europe
AJ252790.1	Europe
AJ252791.1	Europe
AJ252792.1	Europe
AJ252793.1	Europe
AJ252794.1	Europe
AJ252795.1	Europe
AJ252796.1	Europe
AJ252797.1	Europe
AJ252798.1	Europe
AJ252799.1	Europe
AJ252800.1	Europe
AJ252801.1	Europe
AJ252802.1	Europe
AJ252803.1	Europe
AJ252804.1	Europe
AJ252805.1	Europe
AJ252806.1	Europe
AJ252807.1	Europe
AY509826.1	France
AY509827.1	Greece
AY549461.1	Germany
EU856045.1	Italy
EU856046.1	Italy
KU302625.1	Switzerland
KU302630.1	Germany
KU302631.1	Germany
KU302632.1	Germany
KU302635.1	Switzerland

	KU302636.1	Switzerland
	KU302637.1	Germany
	KU302638.1	France
	KU302639.1	Switzerland
	KU302640.1	Germany
	KU302641.1	Germany
	KU302642.1	Switzerland
	Y10446.1	France
Leuciscus idus	AY026397.1	Slovakia
	HM560098.1	Czech Republic
	HM560099.1	Czech Republic
Leuciscus leuciscus	AY509823.1	Canada
	DQ664302.1	Europe
	DQ664303.1	Europe
	DQ664304.1	Europe
	DQ664305.1	Europe
	DQ664306.1	Europe
	HM560100.1	Czech Republic
	HM560101.1	Czech Republic
	Y10449.1	France
Phoxinus phoxinus	EU352213.1	UK
	EU755036.1	Germany
	KX265386.1	South korea
	KX265387.1	South Korea
	KX265388.1	South Korea
	KX265389.1	South Korea
	KX265390.1	South Korea
	KX265391.1	South Korea
	KX265392.1	South Korea
	KX265393.1	South Korea
	KX265394.1	South Korea
	KX265395.1	South Korea
	KX265396.1	South Korea
	KX265397.1	South Korea
	KX265398.1	South Korea
	KX265399.1	South Korea
	KX265400.1	South Korea
	KX265401.1	South Korea
	KX265402.1	South Korea
Pimephales promelas	GQ184519.1	Unknown
	GQ184520.1	Unknown
	GQ184521.1	Unknown
	GQ184522.1	Unknown

	GQ275158.1	Mexico
	GQ275159.1	Mexico
	KU856825.1	Illinois, USA
Pseudorasbora parva	AY952995.1	China
	EU934500.1	China
	EU934501.1	China
	EU934502.1	China
	EU934503.1	China
	EU934504.1	China
	HM117852.1	Europe
	HM117853.1	Europe
	HM117854.1	Europe
	HM117855.1	Europe
	HM117856.1	Europe
	HM117857.1	Europe
	HM117858.1	Europe
	HM117859.1	Europe
	HM117860.1	Europe
	HM117861.1	Europe
	HM117862.1	Europe
	HM117863.1	Europe
	HM117864.1	Europe
	HM117865.1	Europe
	HM117866.1	Europe
	HM117867.1	Europe
	HM117868.1	Europe
	HM117869.1	Europe
	HM117870.1	Europe
	HM117871.1	Europe
	HM117872.1	Europe
	HM117873.1	Europe
	HM117874.1	Europe
	HM117875.1	Europe
	HM117876.1	Europe
	HM117877.1	Europe
	HM117878.1	Europe
	HM117879.1	Europe
	HM117880.1	Europe
	HM117881.1	Europe
	HM117882.1	Europe
	HM117883.1	Europe
	HM117884.1	Europe
	HM117885.1	Europe

	HM117886.1	Europe
	HM117887.1	Europe
	HM117888.1	Europe
	HM117889.1	Europe
	HM117890.1	Europe
	HM117891.1	Europe
	HM117892.1	Europe
	HM117893.1	Europe
	HM117894.1	Europe
	HM117895.1	Europe
	HM117896.1	Europe
	HM117897.1	Europe
	HM117898.1	Europe
	HM117899.1	Europe
	HM117900.1	Europe
	HM117901.1	Europe
	HM224302.1	Japan
	HM560155.1	Turkey
	KP053618.1	South Korea
	Y10453.1	France
Rhodeus sericeus	AB366518.1	China
	DQ396683.1	Russia
	DQ396684.1	Russia
	DQ396685.1	Russia
	DQ396686.1	Russia
	KF410785.1	Russia
	KF410786.1	Russia
	Y10454.1	France
Rutilus rutilus	KF784808.1	Greece
	KF784810.1	Greece
	KF784811.1	Greece
	KF784812.1	Greece
	KF784813.1	Greece
	KF784814.1	Greece
	KF784815.1	Greece
	KF784819.1	Greece
	KF784820.1	Greece
	KF784821.1	Greece
	KF784822.1	Greece
	KF784831.1	Greece
	KF784832.1	Greece
	KF784833.1	Greece
	KF784838.1	Greece

	KF784839.1	Greece
	KF784840.1	Greece
	KF784841.1	Greece
Scardinius erythrophthalmus	AY509835.1	Europe
	AY509836.1	Europe
	AY509837.1	Europe
	AY509838.1	Europe
	AY509839.1	Europe
	AY509840.1	Europe
	AY509841.1	Europe
	AY509842.1	Europe
	AY509843.1	Europe
	AY509844.1	Europe
	AY509845.1	Europe
	AY509846.1	Europe
	AY509847.1	Europe
	AY509848.1	Europe
	EU856057.1	Italy
	HM560171.1	Russia
	Y10444.1	France
Tinca tinca	HM167941.1	Ukraine
	HM167942.1	Bulgaria
	HM167943.1	China
	HM167944.1	UK
	HM167945.1	Romania
	HM167946.1	Turkey
	HM167947.1	Poland
	HM167948.1	Poland
	HM167949.1	Russia
	HM167950.1	France
	HM167951.1	Sweden
	HM167952.1	Germany
	HM167953.1	Czech Republic
	HM167954.1	Germany
	HM167955.1	Iran
	HM167956.1	Iran
	HM167957.1	Iran

Table S4.3: List of non-target species tested using PCR and qPCR to validate assay specificity for *C. carassius*.

Common name	Binomial name	Method
Goldfish	*Carassius auratus*	PCR, qPCR
Common carp	*Cyprinus carpio*	PCR, qPCR
Tench	*Tinca tinca*	PCR, qPCR
Sunbleak	*Leucaspius delineatus*	PCR, qPCR
Common barbel	*Barbus barbus*	qPCR
Topmouth gudgeon	*Pseudorasbora parva*	qPCR
Common rudd	*Scardinius erythrophthalmus*	qPCR
Common roach	*Rutilus rutilus*	qPCR
Stone loach	*Barbatula barbatula*	qPCR
European chub	*Squalius cephalus*	qPCR

Table S4.4: Summary of eDNA amplification by PCR and qPCR for all samples from two ponds.

Sample	PCR amplification (Y/N)	Band strength (copies/µL)	qPCR amplification (Y/N)	DNA copy number (copies/µL)
RAIL1	Y	10-100	Y	78
RAIL2	N	0	N	0
RAIL3	Y	100-1000	Y	306
RAIL4	Y	100-1000	Y	460
RAIL5	Y	10-100	Y	86
MYST1	Y	10-100	Y	11
MYST2	Y	10-100	Y	10
MYST3	Y	10-100	Y	19
MYST4	Y	10-100	Y	15
MYST5	Y	10-100	Y	12

Table S4.5: Summary of model-selection criteria (PPLC and WAIC) for each model containing different covariate combinations fitted to the *C. carassius* eDNA detections. Each model was fit by running the MCMC algorithm for 11,000 iterations and retaining the last 10,000 for estimating posterior summaries. Bold font indicates lowest values of PPLC and WAIC.

Model Number	Model Covariates	PPLC	WAIC
1		120.0547	1.44677854
2	CPUE	119.863	1.445148452
3	pH	119.8931	1.444377085
4	cond	120.2826	1.445161287
5	macrophyte	120.1372	1.447285047
6	CPUE.rep	115.819	1.374592404
7	cond.rep	82.116	1.023717969
8	CPUE, pH	119.9033	1.44721014
9	CPUE, cond	120.1183	1.449389889
10	CPUE, macrophyte	120.0108	1.447009351
11	CPUE, CPUE.rep	115.935	1.379417814
12	CPUE, cond.rep	82.1171	1.025105491
13	pH, cond	119.9968	1.448005204
14	pH, macrophyte	120.0259	1.449985443
15	pH, CPUE.rep	116.0258	1.377743694
16	pH, cond.rep	81.4797	1.013266098
17	cond, macrophyte	120.1835	1.448559093
18	cond, CPUE.rep	116.363	1.380112787
19	cond, cond.rep	83.1228	1.042863594
20	macrophyte, CPUE.rep	115.9689	1.377144366
21	macrophyte, cond.rep	82.9232	1.027835971
22	CPUE.rep, cond.rep	**35.3072**	**0.557895902**
23	CPUE, pH, cond	119.7315	1.452529389
24	CPUE, pH, macrophyte	120.2689	1.448391273
25	CPUE, pH, CPUE.rep	115.857	1.379270198
26	CPUE, pH, cond.rep	81.5833	1.020579022
27	CPUE, cond, macrophyte	120.0777	1.448777854
28	CPUE, cond, CPUE.rep	116.0904	1.381474619
29	CPUE, cond, cond.rep	83.3474	1.049805208
30	CPUE, macrophyte, CPUE.rep	115.6156	1.381069609
31	CPUE, macrophyte, cond.rep	82.6885	1.027532972
32	CPUE, CPUE.rep, cond.rep	35.9762	0.566037955
33	pH, cond, macrophyte	120.0471	1.450851543
34	pH, cond, CPUE.rep	116.049	1.377471375

35	pH, cond, cond.rep	83.7948	1.053803638
36	pH, macrophyte, CPUE.rep	115.9663	1.379459392
37	pH, macrophyte, cond.rep	82.9409	1.02926833
38	pH, CPUE.rep, cond.rep	35.5291	0.550890684
39	cond, macrophyte, CPUE.rep	115.9352	1.376374854
40	cond, macrophyte, cond.rep	83.7313	1.04864047
41	cond, CPUE.rep, cond.rep	35.8065	0.563076678
42	macrophyte, CPUE.rep, cond.rep	36.006	0.563266851
43	CPUE, pH, cond, macrophyte	119.7517	1.454817396
44	CPUE, pH, cond, CPUE.rep	116.0948	1.384483059
45	CPUE, pH, cond, cond.rep	83.7952	1.051519947
46	CPUE, pH, macrophyte, CPUE.rep	116.1947	1.38390534
47	CPUE, pH, macrophyte, cond.rep	83.2494	1.038735101
48	CPUE, pH, CPUE.rep, cond.rep	35.9113	0.563042888
49	CPUE, cond, macrophyte, CPUE.rep	116.1633	1.384283602
50	CPUE, cond, macrophyte, cond.rep	83.8771	1.050351513
51	CPUE, cond, CPUE.rep, cond.rep	36.0218	0.567830368
52	CPUE, macrophyte, CPUE.rep, cond.rep	36.5054	0.571421602
53	pH, cond, macrophyte, CPUE.rep	116.128	1.386414443
54	pH, cond, macrophyte, cond.rep	84.1424	1.054514824
55	pH, cond, CPUE.rep, cond.rep	35.766	0.566525587
56	pH, macrophyte, CPUE.rep, cond.rep	35.9443	0.564220216
57	cond, macrophyte, CPUE.rep, cond.rep	36.1755	0.569288128
58	CPUE, pH, cond, macrophyte, CPUE.rep	116.0153	1.387589582
59	CPUE, pH, cond, macrophyte, cond.rep	84.8881	1.062269558
60	CPUE, pH, cond, CPUE.rep, cond.rep	36.4081	0.573966655
61	CPUE, pH, macrophyte, CPUE.rep, cond.rep	36.3376	0.568899062
62	CPUE, cond, macrophyte, CPUE.rep, cond.rep	36.8728	0.574062559
63	pH, cond, macrophyte, CPUE.rep, cond.rep	36.4053	0.573813761
64	CPUE, pH, cond, macrophyte, CPUE.rep, cond.rep	36.9559	0.573614535

Appendix 4.4: Supplementary figures

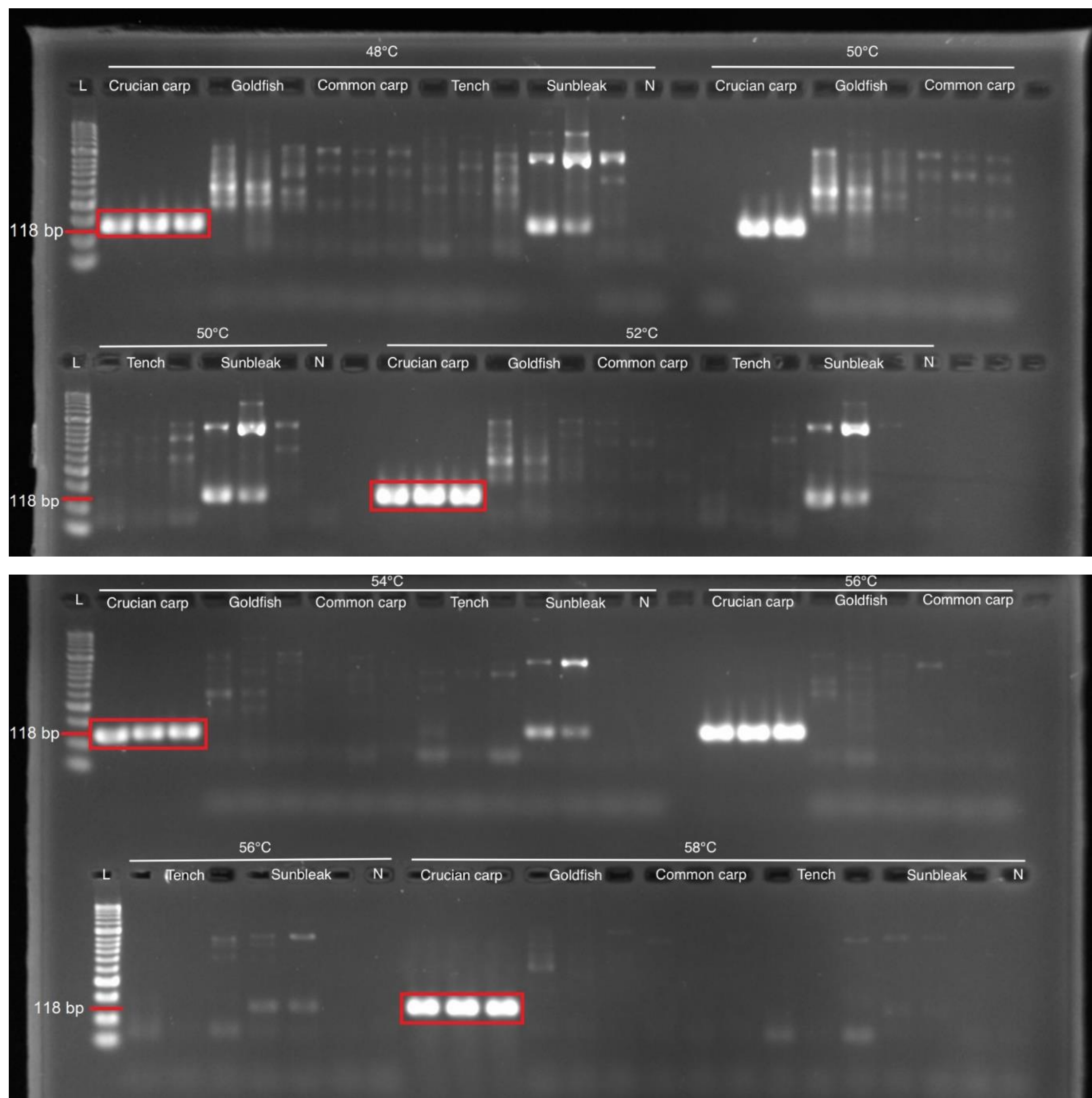


Figure S4.1: PCR products resulting from annealing temperature gradient PCR performed to test specificity of primers for crucian carp (*C. carassius*) against non-target species. Tissue DNA from *C. carassius* was used as a positive control, and three replicates were performed for each fish species. Species name and annealing temperature are given for each set of PCR replicates. Products were run on 2% agarose gels with Hyperladder™ 50bp (Bioline®, UK) molecular weight marker (L). Exemplary bands of expected size (118 bp) are highlighted in red. Specificity was almost achieved at 58 °C, excluding amplification of sunbleak (*Leucaspius delineatus*), thus all future tests were performed at 60 °C.

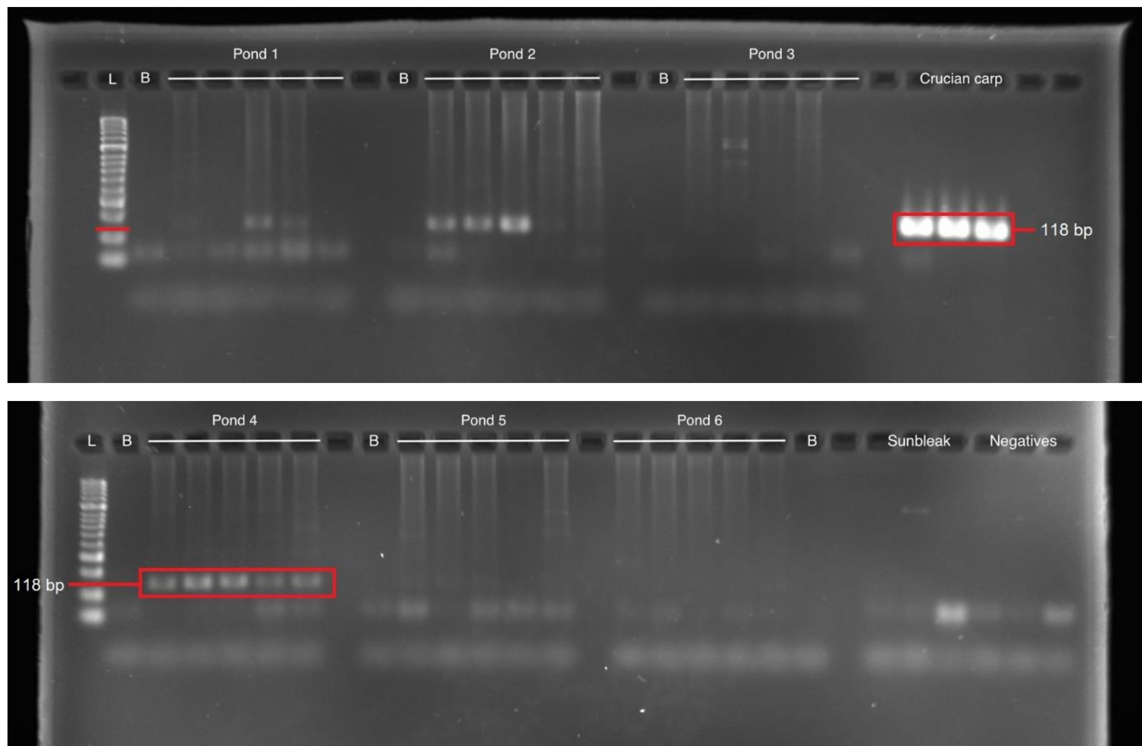


Figure S4.2: PCR products for eDNA samples from ponds recently stocked with crucian carp (*C. carassius*) in Norfolk, eastern England, using species-specific primers at an annealing temperature of 60 °C. Products were run on 2% agarose gels with Hyperladder™ 50bp (Bioline®, UK) molecular weight marker (L). Tissue DNA from *C. carassius* was used as a positive control, and sunbleak (*Leucaspius delineatus*) tissue DNA was also tested to ensure amplification did not occur at the new annealing temperature. Exemplary bands of expected size (118 bp) are highlighted in red. *C. carassius* eDNA was amplified in Ponds 1, 2, 4 and 5, and *L. delineates* DNA did not amplify.

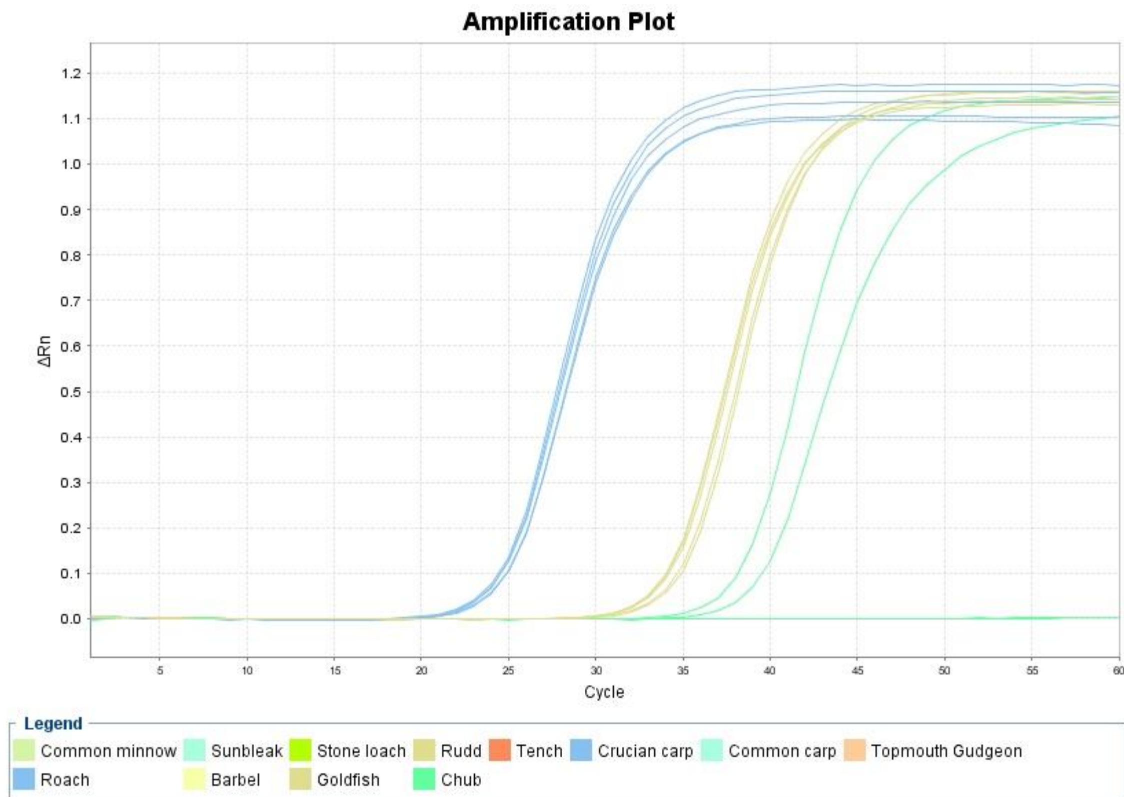


Figure S4.3: qPCR amplification plot for test of primer and probe specificity for crucian carp (*C. carassius*) against 10 non-target fish species. All DNA was standardised to 1 ng/μL. *C. carassius* DNA amplified, but rudd (*Scardinius erythrophthalmus*) and chub (*Squalius cephalus*) also amplified. However, these non-target species produced different amplification curves to *C. carassius* DNA and quantified at lower DNA copy numbers.

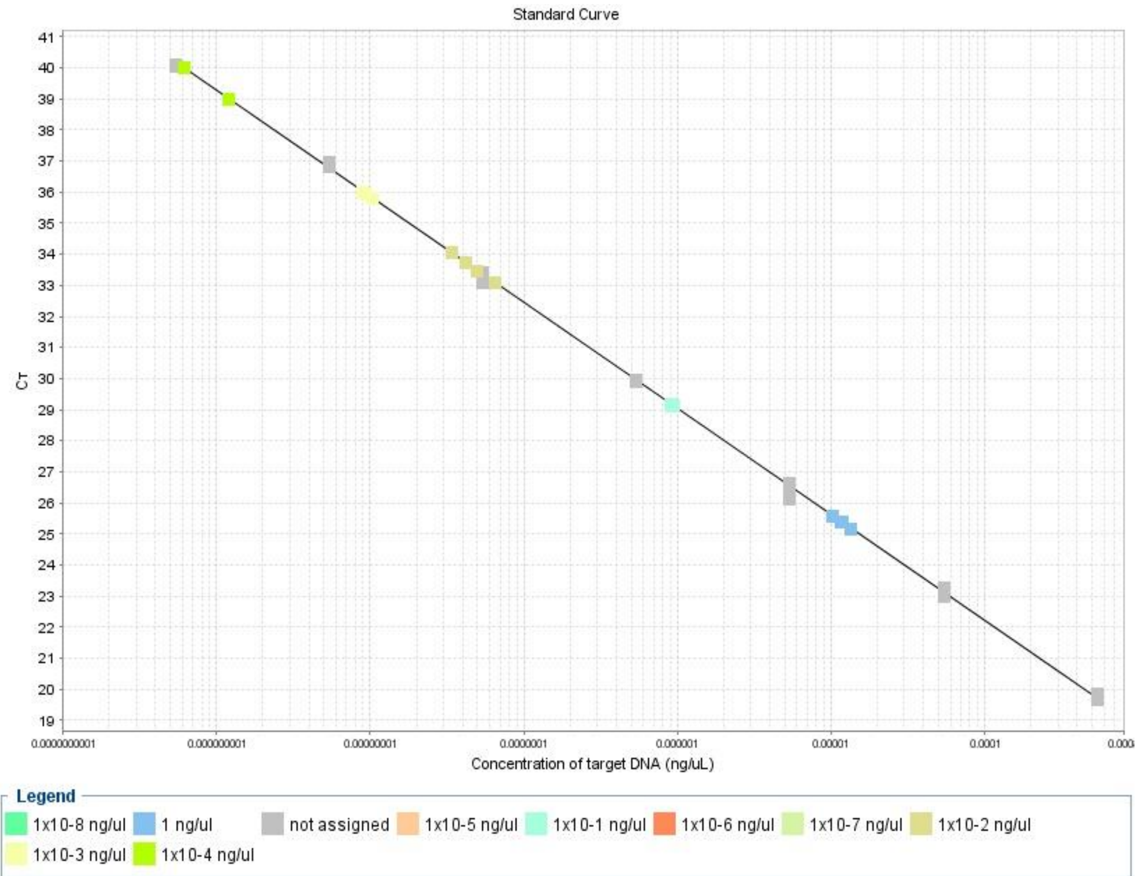


Figure S4.4: qPCR standard curve plot for test of assay sensitivity. All DNA was standardised to 1 ng/µL. qPCR standards (grey points) ranged from 10^6 to 1 copy/µL. The lowest concentration at which *C. carassius* DNA amplified was 0.0001 ng/µL (lime green points).

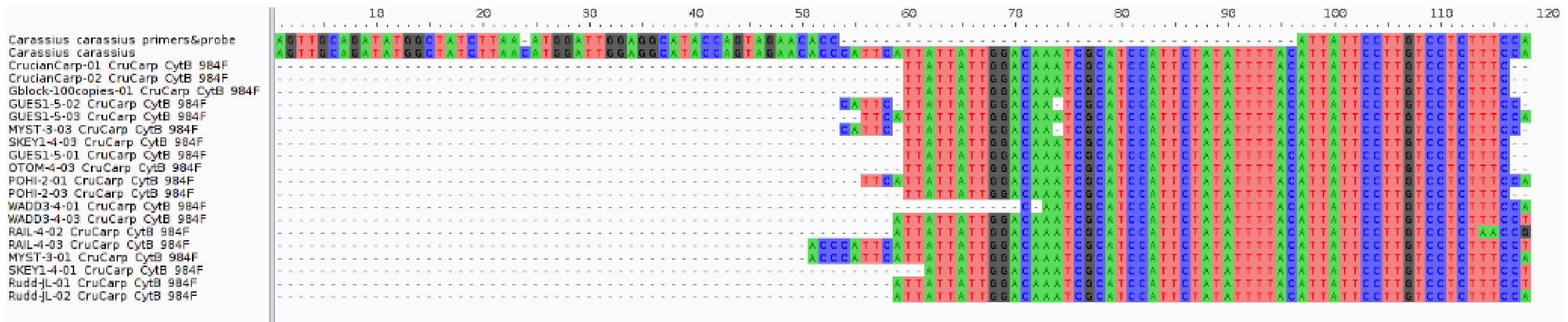


Figure S4.5: Alignment of good quality Sanger sequences obtained for qPCR amplicons from positive controls, representative eDNA samples, and contaminated non-target DNA extracts and full process blank. Species-specific primers and probe for *C. carassius* are given on the first line, followed by the consensus sequence for *C. carassius* used in primer and probe design.

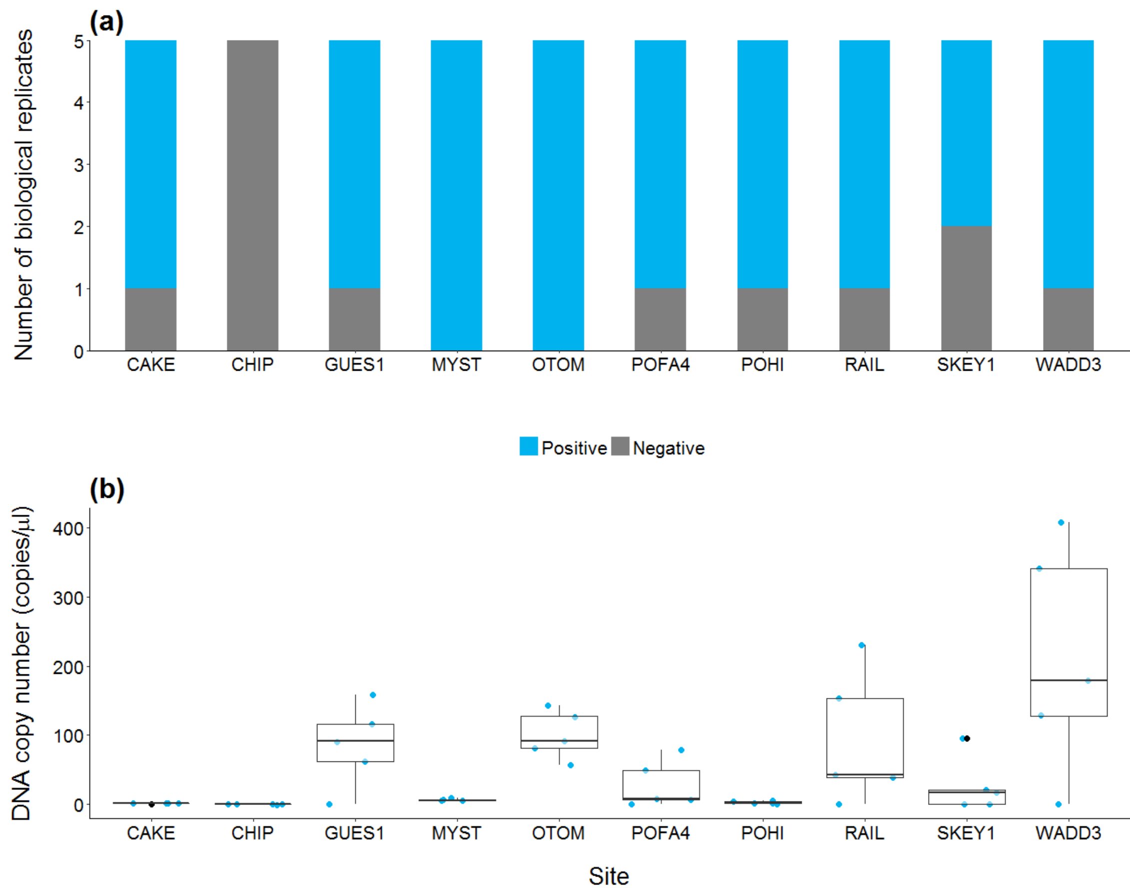


Figure S4.6: Variation amongst eDNA samples for each pond in terms of **(a)** *C. carassius* detection, and **(b)** DNA copy number. The bar chart shows most ponds had four samples that were positive for *C. carassius*, but all samples were negative for the species in one pond. The boxplot represents the distribution of DNA copy number of samples from each pond. The median (line), lower and upper quartiles (lower and upper half of box), and minimum and maximum (whiskers) DNA copy numbers are displayed for each box. DNA copy number was similar in half of the ponds studied but ranged substantially between 0 and 400 copies/µL in others, particularly WADD3.

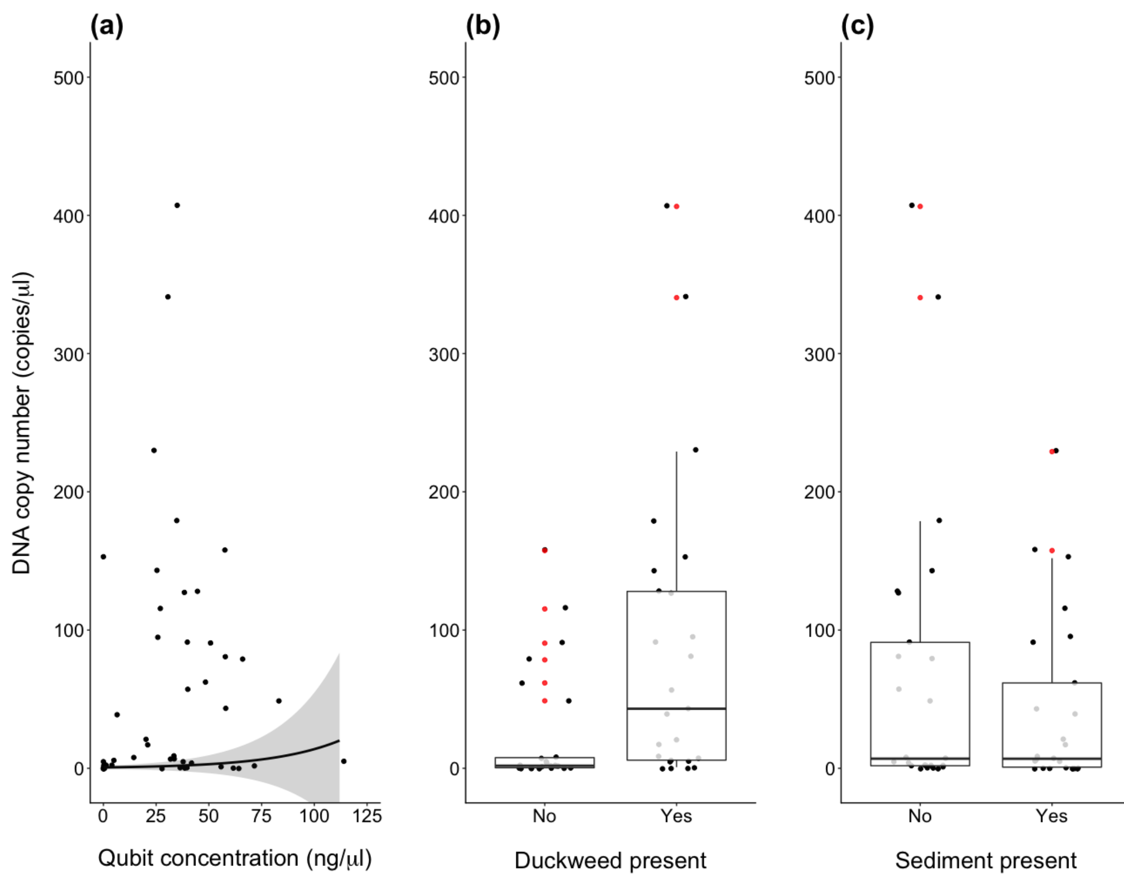


Figure S4.7: Relationship between fixed effects and response variable (DNA copy number) in ponds, as predicted by the hierarchical Poisson GLMM. The 95% CIs, as calculated using the model predictions and standard error for these predictions, are given for each relationship. The observed data (points) are displayed against the predicted relationship (lines/boxes). The scatterplot **(a)** represents the relationship between DNA copy number and Qubit concentration, whereas the boxplots represent **(b)** the distribution of DNA copy number of biological replicates in relation to presence of duckweed (*Lemna* spp.) in ponds and **(c)** the distribution of DNA copy number of biological replicates in relation to presence of sediment in ponds. The median (line), lower and upper quartiles (lower and upper half of box), and minimum and maximum (whiskers) DNA copy numbers are displayed for each box. DNA copy number of biological replicates increased with Qubit concentration **(a)** and where duckweed was present **(b)**, but decreased where sediment was present **(c)**.

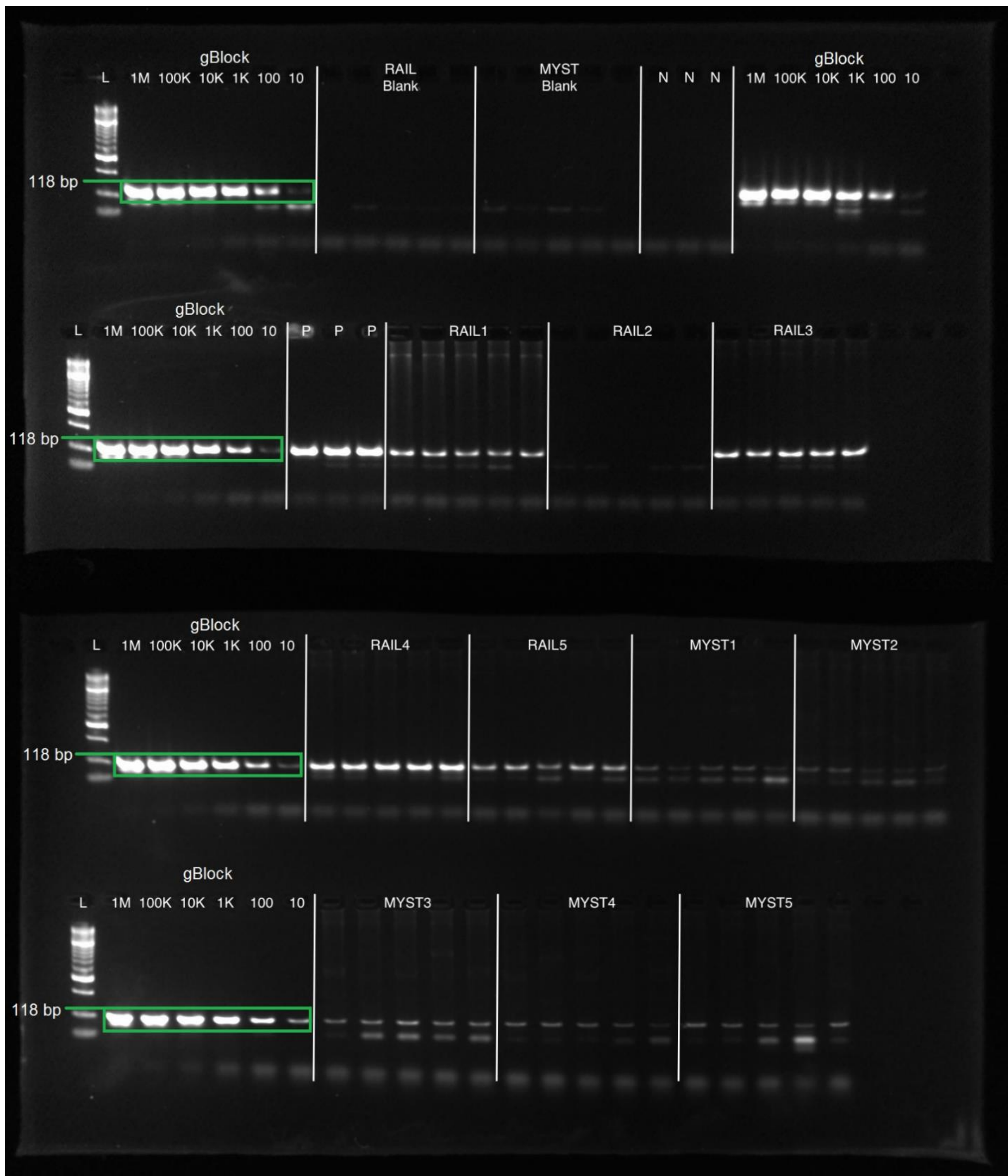


Figure S4.8: PCR products of gBlocks® standards and five eDNA samples from two ponds. Products were run on 2% agarose gels with Hyperladder™ 50bp (Bioline®, UK) molecular weight marker (L). Five replicates were performed for each standard curve point and each eDNA sample. Sample ID is given for each set of replicates, confined by white lines. Exemplary bands of expected size (118 bp) are highlighted in green.

Appendix 4.5: Supplementary references

Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Maechler, M. & Bolker, B.M. (2017) Modeling Zero-Inflated Count Data With glmmTMB. *bioRxiv*, 132753.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J. & Wittwer, C.T. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, **55**, 611–622.

De Ventura, L., Kopp, K., Seppälä, K. & Jokela, J. (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**, 101–112.

Dysthe, J.C., Rodgers, T., Franklin, T.W., Carim, K.J., Young, M.K., McKelvey, K.S., Mock, K.E. & Schwartz, M.K. Repurposing environmental DNA samples—detecting the western pearlshell (*Margaritifera falcata*) as a proof of concept. *Ecology and Evolution*, **8**, 2659–2670.

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.

Fox, J. & Weisberg S. (2011) An R Companion to Applied Regression, Second Edition. Sage, Thousand Oaks, CA.

Hosmer D.W. & Lemeshow S. (2000) Multiple Logistic Regression. In: *Applied Logistic Regression*. John Wiley & Sons, Inc., pp. 31–46.

Lele, S.R., Keim, J.L., & Solymos, P. (2016). ResourceSelection: Resource Selection (Probability) Functions for Use-Availability Data. R package version 0.3-2.

Nathan, L.M., Simmons, M., Wegleitner, B.J., Jerde, C.L. & Mahon, A.R. (2014) Quantifying environmental DNA signals for aquatic invasive species across multiple detection platforms. *Environmental Science & Technology*, **48**, 12800–12806.

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.

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer-Verlag New York, USA. http://ggplot2.org

Wilcox T.M., Carim K.J., McKelvey K.S., Young, M.K. & Schwartz M.K. (2015) The Dual Challenges of Generality and Specificity When Developing Environmental DNA Markers for Species and Subspecies of *Oncorhynchus*. *PLoS ONE*, **10**, e0142008.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A. & Smith, G.M. (2009) *Mixed effects models and extensions in ecology with R*. Springer, New York, USA.

Appendix 5

Appendix 5.1: Supplementary methods

DNA reference database construction

A custom, phylogenetically curated reference database was created for the mitochondrial cytochrome c oxidase subunit I (*COI*) region in UK invertebrate species. A list of recorded UK invertebrate species and their taxonomy was previously constructed by the Centre for Ecology & Hydrology (https://www.ceh.ac.uk/services/coded-macroinvertebrates-list). Database curation for each of the main invertebrate groups (e.g. Coleoptera, Odonata, Mollusca) was performed separately to ease data processing. Reference databases were constructed using the ReproPhylo environment (Szitenberg *et al.*, 2015) in a Jupyter notebook (Kluyver *et al.*, 2016). Jupyter notebooks detailing the processing steps for each data subset are deposited in a dedicated GitHub repository for this chapter (https://github.com/lrharper1/LRHarper_PhDThesis_Chapter5) which has been permanently archived (https://doi.org/10.5281/zenodo.2634240). We used a BioPython script to perform a GenBank search based on the species lists and downloaded all available *COI* sequences for specified species. Where there were no records on GenBank for a UK species, the database was supplemented with downloaded sequences belonging to European sister species in the same genus. Species that had no *COI* records on Genbank are listed in the archived GitHub repository (https://doi.org/10.5281/zenodo.2634240). Importantly, a reference database could not be constructed for Diptera due to problems encountered with taxonomy.

Redundant sequences were removed by clustering at 100% similarity using vsearch v1.1 (Rognes *et al.*, 2016). Only sequences longer than 500 bp were processed to increase alignment robustness to large gaps. Sequences were aligned using MUSCLE (Edgar, 2004). Alignments were trimmed using trimAl (Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009), following which maximum likelihood trees were inferred with RAxML 8.0.2 (Stamatakis, 2006) using the GTR+gamma model of substitutions. The complete alignments were then processed using SATIVA (Kozlov *et al.*, 2016) for automated identification of 'mislabelled' sequences which could cause conflict in downstream analyses. Putatively mislabelled sequences were removed, whereupon alignment and phylogenetic tree construction were repeated for manual investigation of

sequences. The resultant databases (i.e. curated non-redundant reference databases) contained sequences from: 412/423 Coleoptera species, 54/59 Odonata species, 83/92 Ephemeroptera/Plecoptera/Nemoptera/Megaloptera species, 187/206 Trichoptera/Lepidoptera species, 53/114 Hemiptera/Hymenoptera species, 154/388 Crustacea species, 78/111 Mollusca species, 333/333 Arachnida species, and 129/152 Annelida species. These databases were used for *in silico* validation of primers.

The invertebrate databases were supplemented by Sanger sequences obtained from tissue of a two-spotted assassin bug (*Platymeris biguttatus*) housed at the University of Hull. DNA from this species was used as our PCR positive control. *P. biguttatus* DNA was extracted from tissue samples using a DNeasy Blood & Tissue Kit® (Qiagen®, Hilden, Germany). Reference sequences were generated using the standard *COI* primers for DNA barcoding of invertebrates (Folmer *et al.*, 1994). PCR reactions were performed in 25 µL volumes containing: 12.5 µL of MyTaq™ HS Red Mix (Bioline®, UK), 1 µL (final concentration - 0.04 µM) of forward and reverse primer (Integrated DNA Technologies, Belgium), 8.5µL of molecular grade sterile water (Fisher Scientific UK Ltd, UK) and 2 µL DNA template. PCRs were performed on an Applied Biosystems® Veriti Thermal Cycler (Fisher Scientific UK Ltd, UK) with the following profile: 94 °C for 3 min, 37 cycles of 94 °C for 30 sec, 52 °C for 60 sec and 72 °C for 90 sec, followed by a final elongation step at 72 °C for 10 min. Purified PCR products were Sanger sequenced directly (Macrogen Europe, Amsterdam, Netherlands) in both directions using the PCR primers. Sequences were edited using CodonCode Aligner (CodonCode Corporation, MA, USA). The complete invertebrate reference database compiled in GenBank format has been deposited in the GitHub repository for this chapter.

Primer validation

Invertebrates from bulk tissue DNA and environmental DNA (eDNA) samples were amplified with published *COI* primers mICOIintF (Leray *et al.*, 2013) and jgHCO2198 (Geller *et al.*, 2013). These primers were validated for the present study *in silico* using ecoPCR software (Ficetola *et al.*, 2010) against the custom, phylogenetically curated reference database for UK invertebrates. Parameters were set to allow a fragment size of 250-350 bp and maximum of three mismatches between the primer pair and each sequence in the reference database. Primers were also validated *in vitro* for tissue DNA extracted from 38 invertebrate species that represented 38 families and 10 major groups.

During *in vitro* testing, the chosen primers were compared to two other published primer sets for macroinvertebrates: BF2/BR2 (Elbrecht & Leese, 2017) and fwhF1/fwhR1 (Vamos, Elbrecht & Leese, 2017). Primer validation tests were performed at the University of Hull in a separate laboratory situated on a different floor to the dedicated eDNA laboratory. All PCR reactions were performed in 25 μL volumes containing: 12.5 μL of MyTaq™ HS Red Mix (Bioline®, UK), 1 μL (final concentration - 0.04 μM) of forward and reverse primer (Integrated DNA Technologies, Belgium), 8.5μL of molecular grade sterile water (Fisher Scientific UK Ltd, UK) and 2 μL DNA template. Thermocycling conditions were kept as consistent as possible across different primer sets tested, bar annealing temperature. PCRs were performed on an Applied Biosystems® Veriti Thermal Cycler (Fisher Scientific UK Ltd, UK) with the following profile: 95 °C for 3 min, 35 cycles of 95 °C for 30 sec, x °C for 30 sec, and 72 °C for 60 sec, followed by a final elongation step at 72 °C for 10 min. Annealing temperatures for mICOIintF/jgHCO2198, BF2/BR2, and fwhF1/fwhR1 were 51 °C, 50 °C, and 52 °C respectively. Molecular grade sterile water (Fisher Scientific UK Ltd, UK) substituted template DNA for PCR negative controls.

Metabarcoding workflow

A two-step PCR protocol was performed on bulk tissue and eDNA samples at the University of Hull. For bulk tissue samples, PCR reactions were set up in an ultraviolet and bleach sterilised laminar flow hood in a laboratory for analysis of tissue DNA with separate rooms for pre-PCR and post-PCR processes. eDNA samples were processed in the dedicated eDNA facility at the University of Hull with separate rooms for filtration, DNA extraction and PCR preparation of sensitive environmental samples. PCR reactions of eDNA samples were also set up in a UV and bleach sterilized laminar flow hood. For both sample types, eight-strip PCR tubes with individually attached lids were used instead of 96-well plates to minimise cross-contamination risk between samples (Port *et al.*, 2016). PCR positive and negative controls were included on each PCR run (typically two positive and negative controls on each 96-well run), to screen for sources of potential contamination. The DNA used for the PCR positive control (tissue DNA $N = 9$, eDNA $N = 11$) was *P. biguttatus,* as this is an exotic, terrestrial species not found in UK freshwater habitats whose DNA had not been handled in our laboratory prior to this study.

The negative controls (tissue DNA N = 9, eDNA N = 11) substituted molecular grade sterile water (Fisher Scientific UK Ltd, UK) for template DNA.

During the first PCR, the target region was amplified using the primers described above, including adapters (Illumina, 2011). First PCR reactions were performed in triplicate in a final volume of 20 μL, using 2 μL of DNA extract as a template. The amplification mixture contained 10 μL of MyTaq™ HS Red Mix (Bioline®, UK), 1 μL (final concentration - 0.5 μM) of forward and reverse primer (Integrated DNA Technologies, Belgium) and 6 μL of molecular grade sterile water (Fisher Scientific UK Ltd, UK). PCRs were performed on an Applied Biosystems® Veriti Thermal Cycler (Fisher Scientific UK Ltd, UK) and PCR conditions for the first component of the two-step protocol consisted of: an incubation step at 95 °C for 3 min, followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 47 °C for 30 s, and extension at 72 °C for 1 min, with final extension at 72 °C for 10 min. PCR products were stored at 4 °C until fragment size was verified by visualising 2 μL of selected PCR products on 2% agarose gels (80 mL 1x Sodium Borate buffer, 1.6 g agarose powder). Gels were stained with ethidium bromide or GelRed® and imaged using Image Lab Software (Bio-Rad Laboratories Ltd, UK). A PCR product was deemed positive where there was an amplification band on the gel that was of the expected size (400-500 bp). PCR replicates for each eDNA sample were pooled in preparation for the addition of Illumina indexes in the second PCR, which resulted in 60 μL of PCR product for each sample. PCR replicates for bulk tissue samples, PCR positive controls, and PCR negative controls were not pooled to allow individual purification and sequencing. All PCR products were purified to remove excess primer using magnetic bead clean-up. Mag-Bind® RxnPure Plus beads (Omega Bio-tek, GA, USA) were used while following a double size selection protocol from Bronner *et al*. (2009). Magnetic bead ratios of 0.5x and 0.12x to 20 μL of first PCR product were used. Eluted DNA (15 μL) was stored at -20 °C until the second PCR could be performed.

In the second PCR, Multiplex Identification (MID) tags (unique 8-nucleotide sequences) and Illumina MiSeq adapter sequences were bound to the amplified product. For each second PCR run, 96 unique tag combinations were created by combining eight unique forward tags with 12 unique reverse tags or vice versa (Kitson *et al.*, 2019). A total of 384 unique tag combinations were achieved, allowing samples to be distinguished during bioinformatics analysis. Second step PCR reactions were performed in eight-strip PCR tubes with individually attached lids in a final volume of 50 μL, using 5 μL of

purified DNA from the first PCR product as a template. The amplification mixture contained 25 µL of MyTaq™ HS Red Mix (Bioline®, UK), 5 µL (final concentration - 0.4 µM) of tagged primer mix (Integrated DNA Technologies, Belgium) and 15 µL of molecular grade sterile water (Fisher Scientific UK Ltd, UK). PCR was performed on an Applied Biosystems® Veriti Thermal Cycler (Fisher Scientific UK Ltd, UK) with the following profile: denaturation at 95 °C for 3 min, followed by 8 cycles of denaturation at 95 °C for 15 s, annealing at 55 °C for 30 s, and extension at 72 °C for 30 s, with final extension at 72 °C for 10 min. PCR products were stored at 4 °C before they were visualised on 2% agarose gels (80 mL 1x Sodium Borate buffer, 1.6 g agarose powder) using 2 µL PCR product. Gels were stained with ethidium bromide or GelRed® and imaged using Image Lab Software (Bio-Rad Laboratories Ltd, UK). Again, PCR products were deemed positive where there was an amplification band on the gel that was of the expected size (500-600 bp). Second PCR products (25 µL) were pooled according to PCR run, and the pooled PCR products purified to remove excess primer using magnetic bead clean-up. Mag-Bind® RxnPure Plus beads (Omega Bio-tek, GA, USA) were used while following a double size selection protocol from Bronner *et al*. (2009). Magnetic bead ratios of 0.5x and 0.12x to 200 µL of pooled PCR product were used. Eluted DNA (30 µL) was stored at -20 °C until library quality control.

The pooled PCR products were quantified on a Qubit™ 3.0 fluorometer using a Qubit™ dsDNA HS Assay Kit (Invitrogen, UK) and pooled proportional to sample size and concentration. The pooled libraries were then quantified on a Qubit™ 3.0 fluorometer using a Qubit™ dsDNA HS Assay Kit (Invitrogen, UK). Based on Qubit™ concentration, the libraries were diluted to 6 nM for quantification by real-time quantitative PCR (qPCR) using the NEBNext® Library Quant Kit for Illumina® (New England Biolabs® Inc., MA, USA). The libraries were also checked using an Agilent 2200 TapeStation and High Sensitivity D1000 ScreenTape (Agilent Technologies, CA, USA) to verify secondary product had been removed successfully and a fragment of the expected size (531 bp) remained. The bulk tissue library was sequenced at 15 pM with 10% PhiX Control and eDNA library sequenced at 8pM with 20% PhiX Control on an Illumina MiSeq® using 2 x 300 bp V3 chemistry (Illumina Inc., CA, USA).

Illumina® data was converted from raw sequences to taxonomic assignment using a custom pipeline for reproducible analysis of metabarcoding data: metaBEAT (metaBarcoding and eDNA Analysis Tool) v0.97.11 (https://github.com/HullUni-bioinformatics/metaBEAT). Raw reads were quality trimmed using Trimmomatic v0.32

(Bolger, Lohse & Usadel, 2014), both from the read ends (minimum per base phred score Q30), as well as across sliding windows (window size 5bp; minimum average phred score Q30). Reads were clipped to a length of 200 bp and reads shorter than 200 bp after quality trimming were discarded. To reliably exclude adapters and PCR primers, the first 26 bp of all remaining reads were also removed. Sequence pairs were merged into single high quality reads using FLASH v1.2.11 (Magoč & Salzberg, 2011), if a minimum of 10 bp overlap with a maximum of 10% mismatch was detected between pairs. For reads that were not successfully merged, only forward reads were kept. To reflect our expectations with respect to fragment size, a final length filter was applied and only sequences of length 313 bp were retained. These were screened for chimeric sequences against our custom reference database using the uchime algorithm (Edgar *et al.*, 2011), as implemented in vsearch v1.1.0 (Rognes *et al.*, 2016). Redundant sequences were removed by clustering at 97% identity ('--cluster_fast' option) in vsearch v1.1.0 (Rognes *et al.*, 2016). Clusters represented by less than three sequences were considered sequencing error and omitted from further analyses. Non-redundant sets of query sequences were then compared against our custom reference database using BLAST (Zhang *et al.*, 2000). For any query matching with at least 90% identity to a reference sequence across more than 80% of its length, putative taxonomic identity was assigned using a lowest common ancestor (LCA) approach based on the top 10% BLAST matches. Sequences that could not be assigned (non-target sequences) were subjected to a separate BLAST search against the complete NCBI nucleotide (nt) database at 90% identity to determine the source via LCA as described above. To ensure reproducibility of analyses, the described workflow has been deposited in the GitHub repository.

Appendix 5.2: Supplementary results

The *in silico* analysis indicated poor taxonomic coverage and resolution of the *COI* primers, where only 9.24% of target invertebrate species amplified. A small range of UK invertebrate taxa were amplified, with fragment length ranging from 307-313 bp. The primers amplified 18/78 Coleoptera species, 16/54 Odonata species, 9/83 Ephemeroptera/Plecoptera/Nemoptera/Megaloptera species, 8/187 Trichoptera/Lepidoptera species, 4/53 Hemiptera/Hymenoptera species, 20/154 Crustacea species, 18/78 Mollusca species, 10/333 Arachnida species, and 29/129 Annelida species (Fig. S5.1). However, an important caveat of these results is available reference sequences on GenBank. The majority of invertebrate *COI* sequences on GenBank were generated using the Folmer primers, LCO1490 and HCO2198 (Folmer *et al.*, 1994). After Sanger sequencing, primer regions are often removed due to low quality sequence produced at the start of sequencing. Therefore, primer sequences are typically not included in invertebrate reference sequences uploaded to GenBank. Our forward metabarcoding primer mICOIintF (Leray *et al.*, 2013) lies within the 658 bp fragment amplified by the Folmer primers. However, our reverse metabarcoding primer jgHCO2198, a modified version of HCO2198 (Geller *et al.*, 2013), lies outwith this fragment. Consequently, ecoPCR is unable to find any match between the reverse primer (jgHCO2198) and sequences in the invertebrate reference databases, causing *in silico* amplification failure. During *in vitro* tests, bands were observed by agarose gel electrophoresis for all invertebrate tissue tested (representing 38 species and 38 families), and no bands were observed in PCR negative controls (Fig. S5.2).

Appendix 5.3: Supplementary tables

Table S5.1: Summary of number of invertebrate species detected by each method across each invertebrate group.

Group	Sweep-netting and microscopy	DNA metabarcoding	eDNA metabarcoding	Total
Annelida	0	18	17	35
Arachnida	0	1	5	6
Bryozoa	0	1	2	3
Cnidaria	0	0	2	2
Coleoptera	21	27	5	53
Collembola	0	1	1	2
Crustacea	4	6	25	35
Diptera	0	26	27	53
Ephemeroptera	1	2	2	5
Gastrotricha	0	1	9	10
Hemiptera	19	11	7	37
Hirudinea	7	7	2	16
Hymenoptera	0	0	1	1
Lepidoptera	0	0	1	1
Megaloptera	1	1	0	2
Mollusca	24	17	13	54
Nematoda	0	0	1	1
Odonata	11	11	4	26
Platyhelminthes	0	0	2	2
Psocoptera	0	0	1	1
Rotifera	0	7	27	34
Tardigrada	0	1	1	2

Thysanoptera	0	0	1	1
Trichoptera	3	3	4	10
Total	**91**	**141**	**160**	**392**

Table S5.2: Summary of number of invertebrate families detected by each method across each invertebrate group.

Group	Sweep-netting and microscopy	DNA metabarcoding	eDNA metabarcoding	Total
Annelida	0	3	5	8
Arachnida	0	2	8	10
Bryozoa	0	1	2	3
Cnidaria	0	0	1	1
Coleoptera	7	7	4	18
Collembola	0	1	4	5
Crustacea	3	8	12	23
Diptera	6	8	13	27
Ephemeroptera	1	1	2	4
Gastrotricha	0	1	2	3
Hemiptera	6	4	5	15
Hirudinea	2	2	1	5
Hymenoptera	0	0	3	3
Lepidoptera	0	0	1	1
Megaloptera	1	1	0	2
Mollusca	7	8	8	23
Nematoda	0	0	1	1
Odonata	3	3	2	8
Platyhelminthes	0	2	2	4
Psocoptera	0	0	1	1
Rotifera	0	2	10	12
Tardigrada	0	1	2	3
Thysanoptera	0	0	1	1
Trichoptera	2	2	2	6

Total	**38**	**57**	**92**	**187**

Appendix 5.4: Supplementary figures

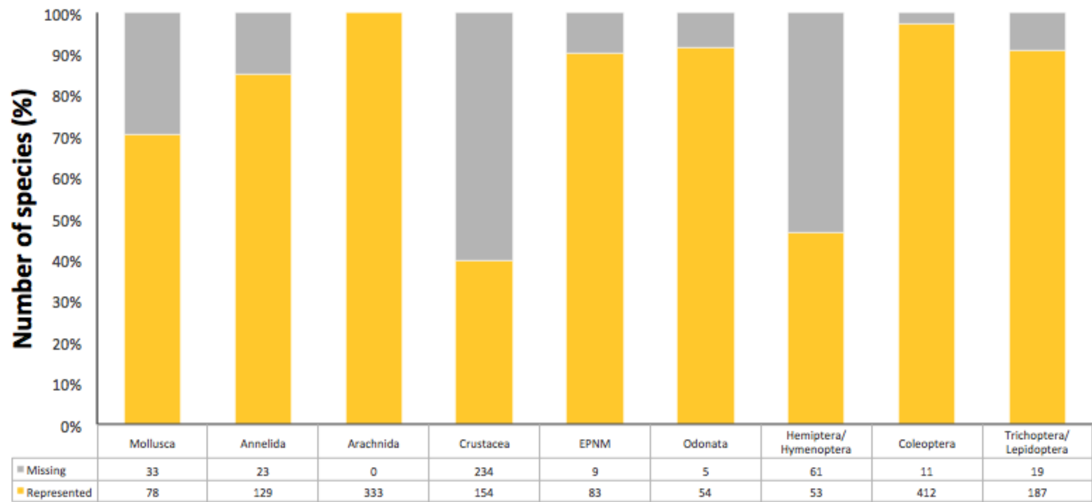


Figure S5.1: Barplot summarising the number and proportion of species with and without reference sequences from GenBank for each custom invertebrate database.

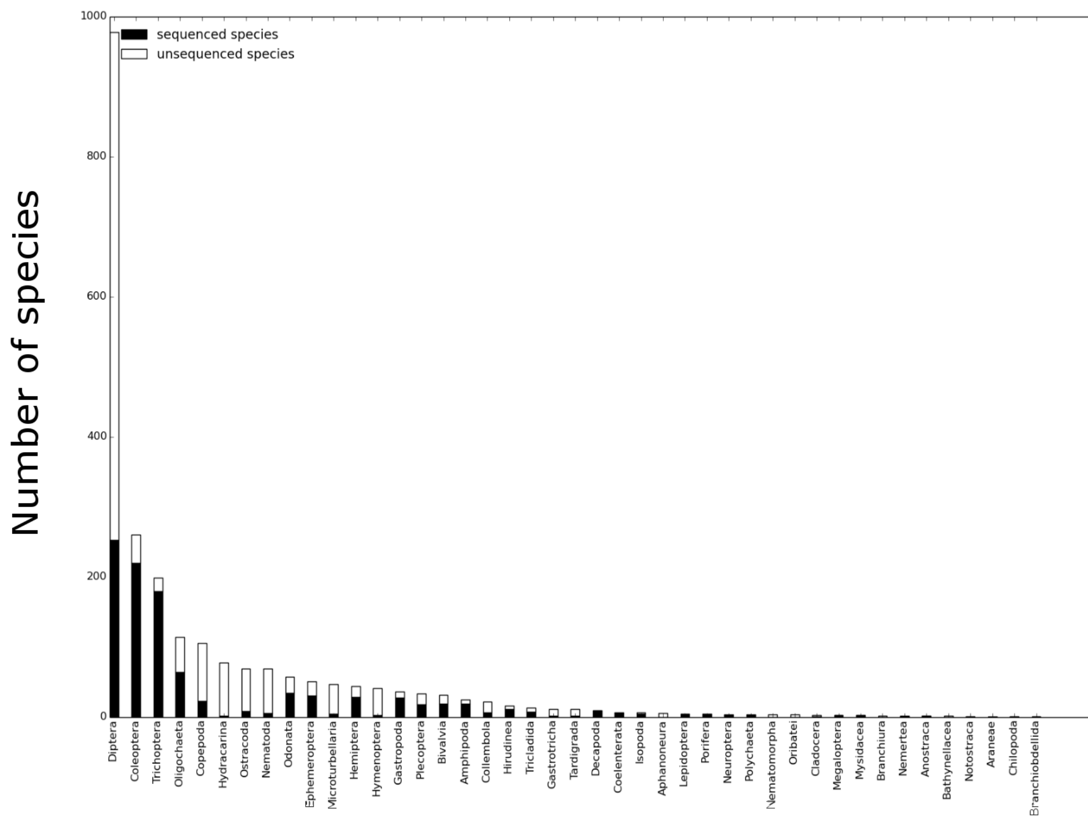
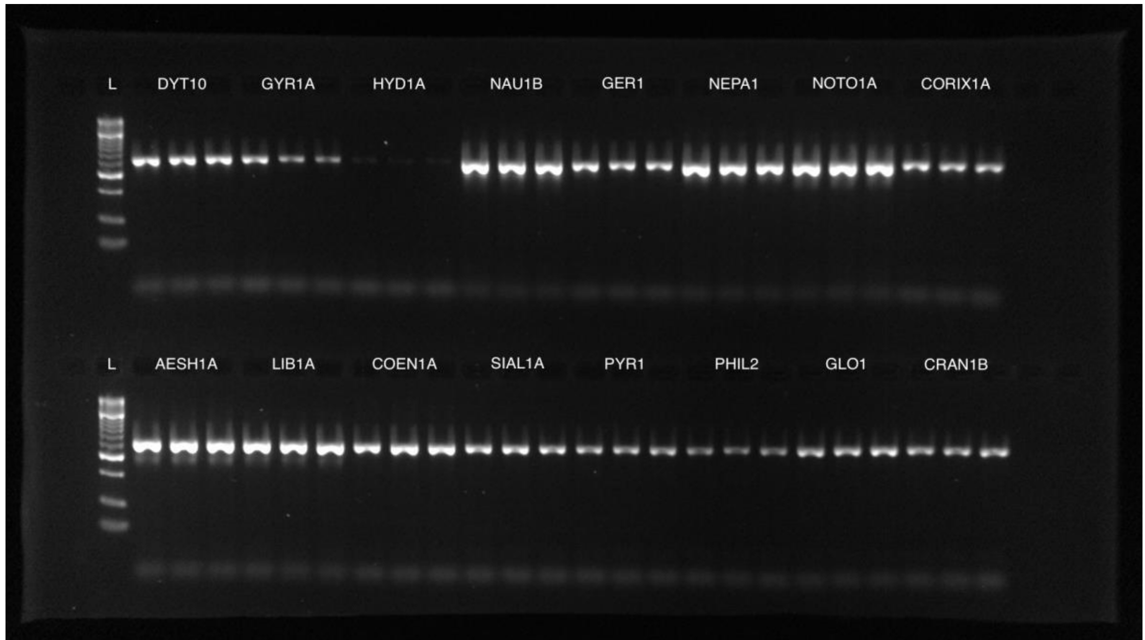
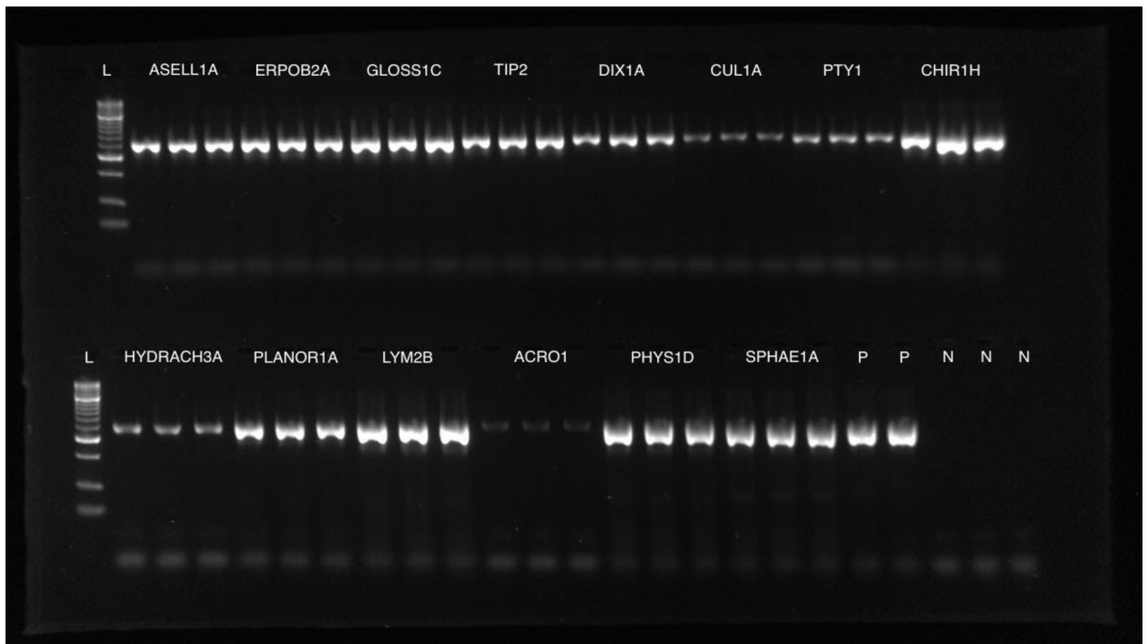


Figure S5.2: Barplot summarising the number of species with and without records on GenBank according to invertebrate groups.


L DYT10 GYR1A HYD1A NAU1B GER1 NEPA1 NOTO1A CORIX1A
L AESH1A LIB1A COEN1A SIAL1A PYR1 PHIL2 GLO1 CRAN1B


L ASELL1A ERPOB2A GLOSS1C TIP2 DIX1A CUL1A PTY1 CHIR1H
L HYDRACH3A PLANOR1A LYM2B ACRO1 PHYS1D SPHAE1A P P N N N

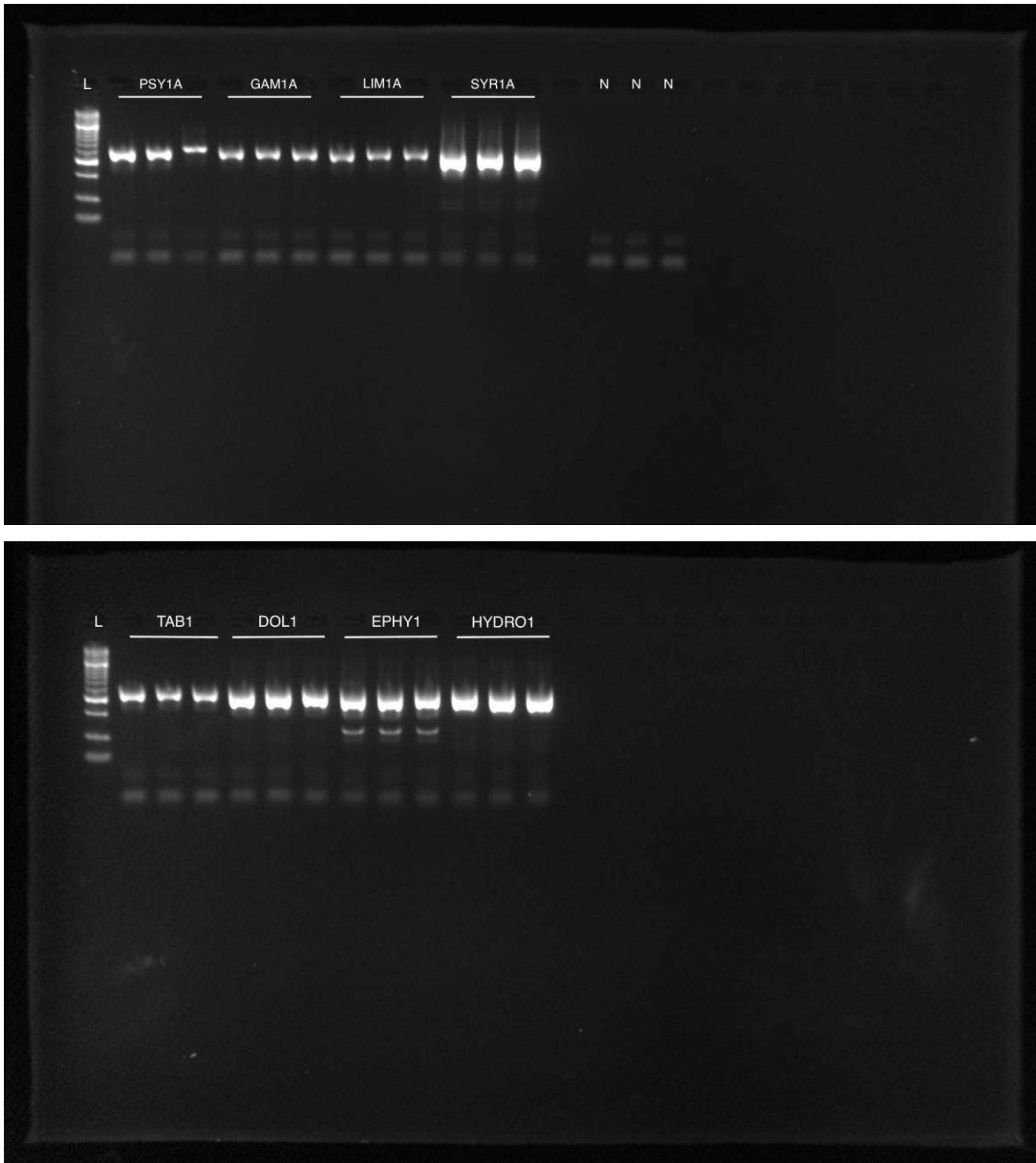


Figure S5.3: Gel images showing results of *in vitro* primer validation for primers mICOIintF and jgHCO2198. PCR products were run on 2% agarose gels with Hyperladder™ 50bp (Bioline®, UK) molecular weight marker (L). Tissue from the two-spotted assassin bug (*Platymeris biguttatus*) was used as the positive control.

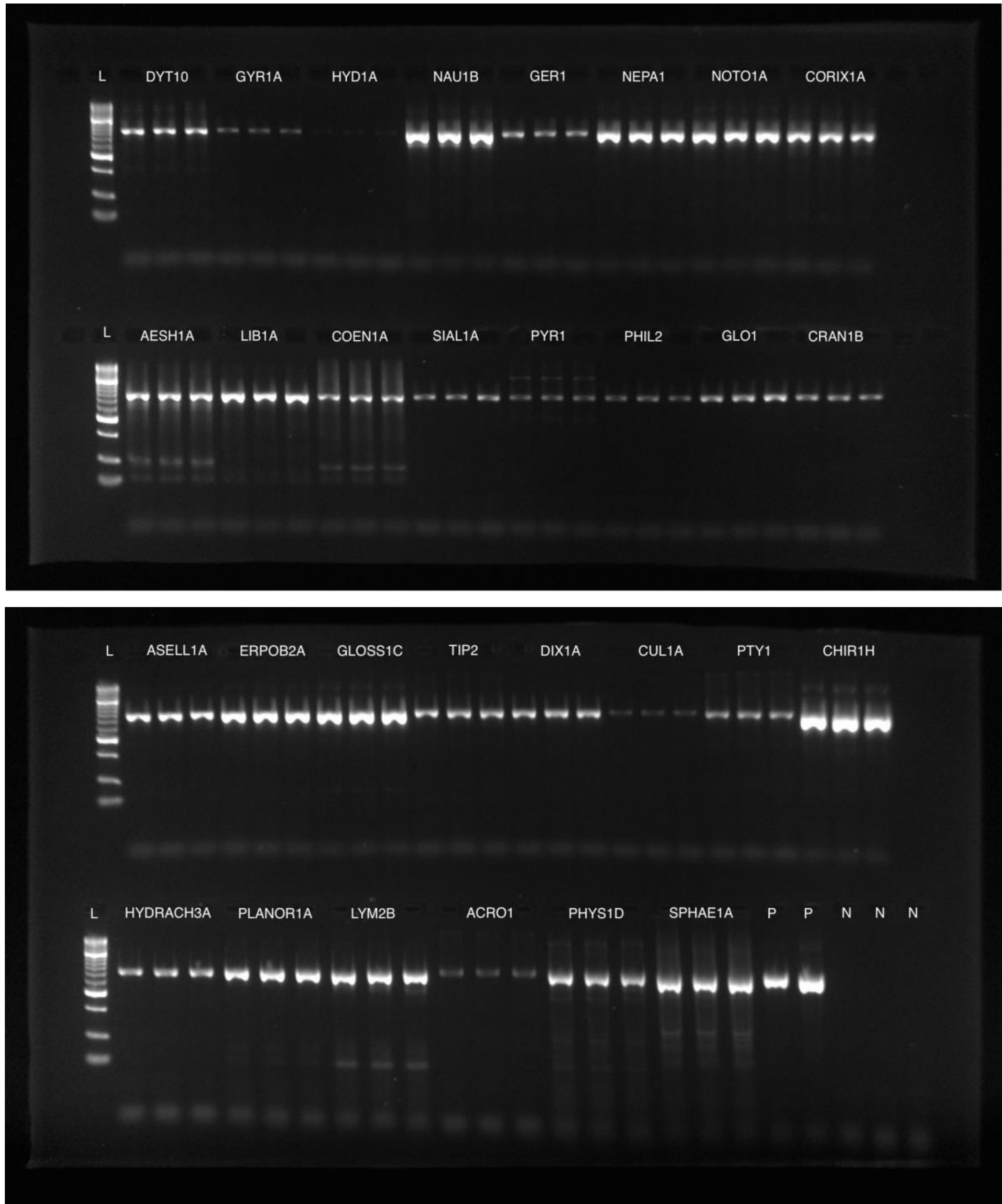


Figure S5.4: Gel images showing results of PCR for primers BF2 and BR2 (Elbrecht & Leese, 2017). PCR products were run on 2% agarose gels with Hyperladder™ 50bp (Bioline®, UK) molecular weight marker (L). Tissue from the two-spotted assassin bug (*Platymeris biguttatus*) was used as the positive control.

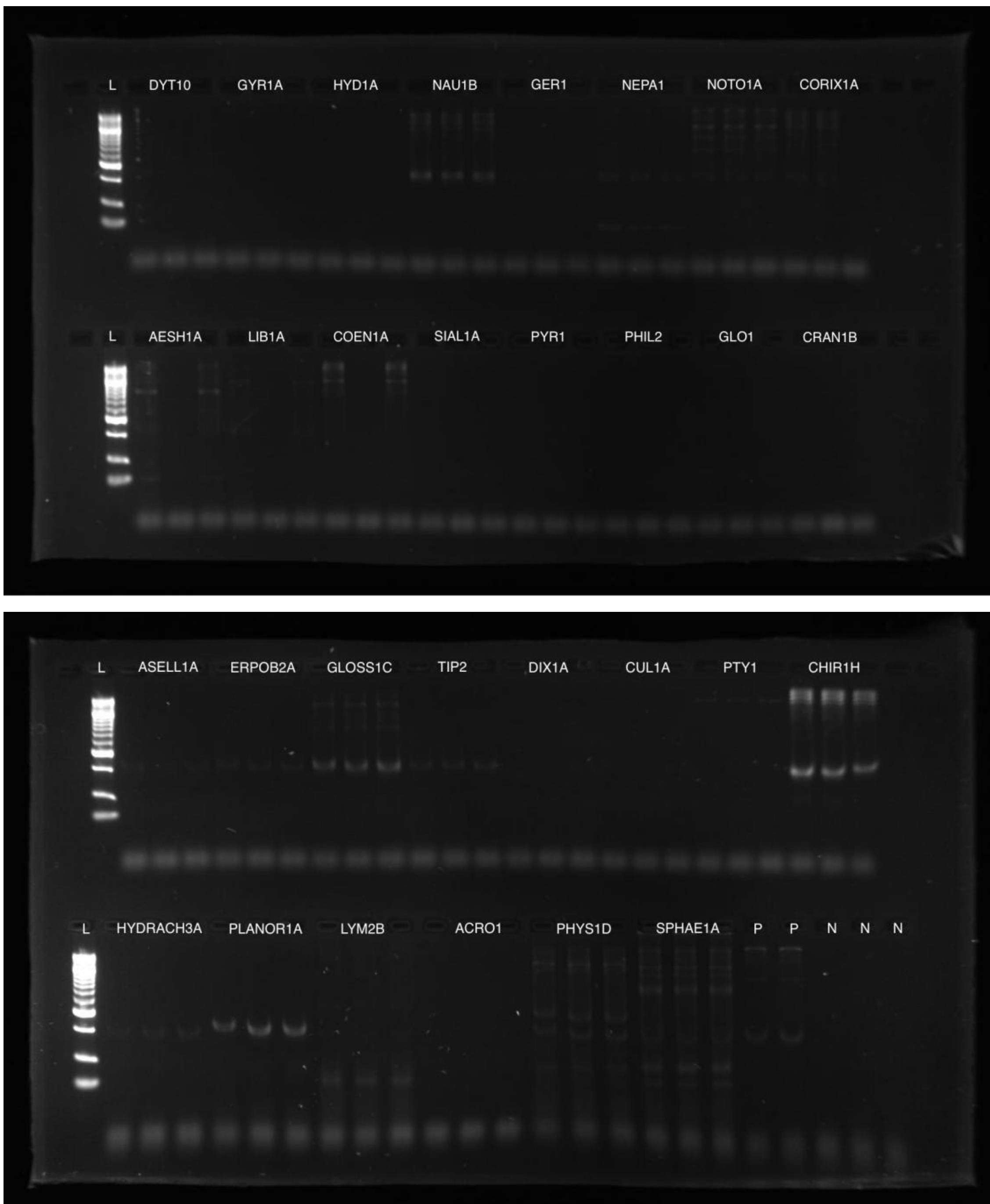


Figure S5.5: Gel images showing results of PCR for primers fwhF1 and fwhR1 (Vamos *et al.*, 2017). PCR products were run on 2% agarose gels with Hyperladder™ 50bp (Bioline®, UK) molecular weight marker (L). Tissue from the two-spotted assassin bug (*Platymeris biguttatus*) was used as the positive control.

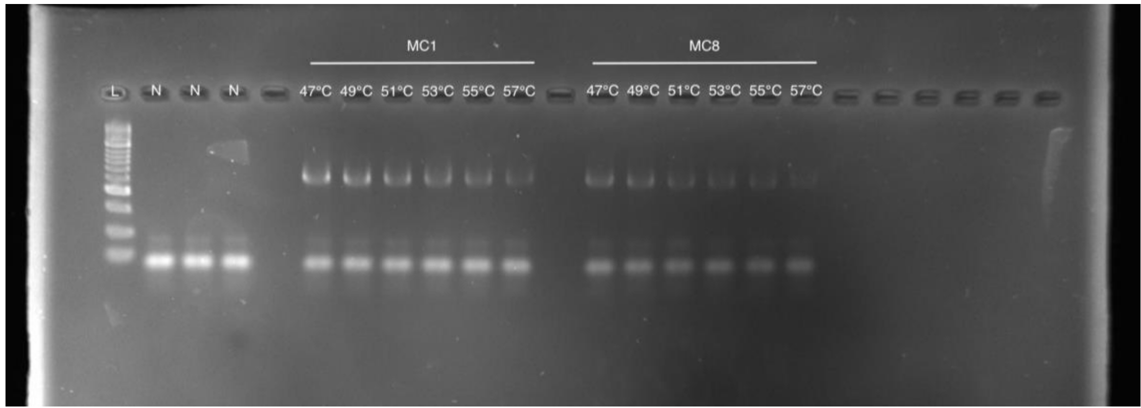


Figure S5.6: Gel image showing results of annealing temperature gradient PCR for primers mICOIintF and jgHCO2198. PCR products were run on 2% agarose gels with Hyperladder™ 50bp (Bioline®, UK) molecular weight marker (L).

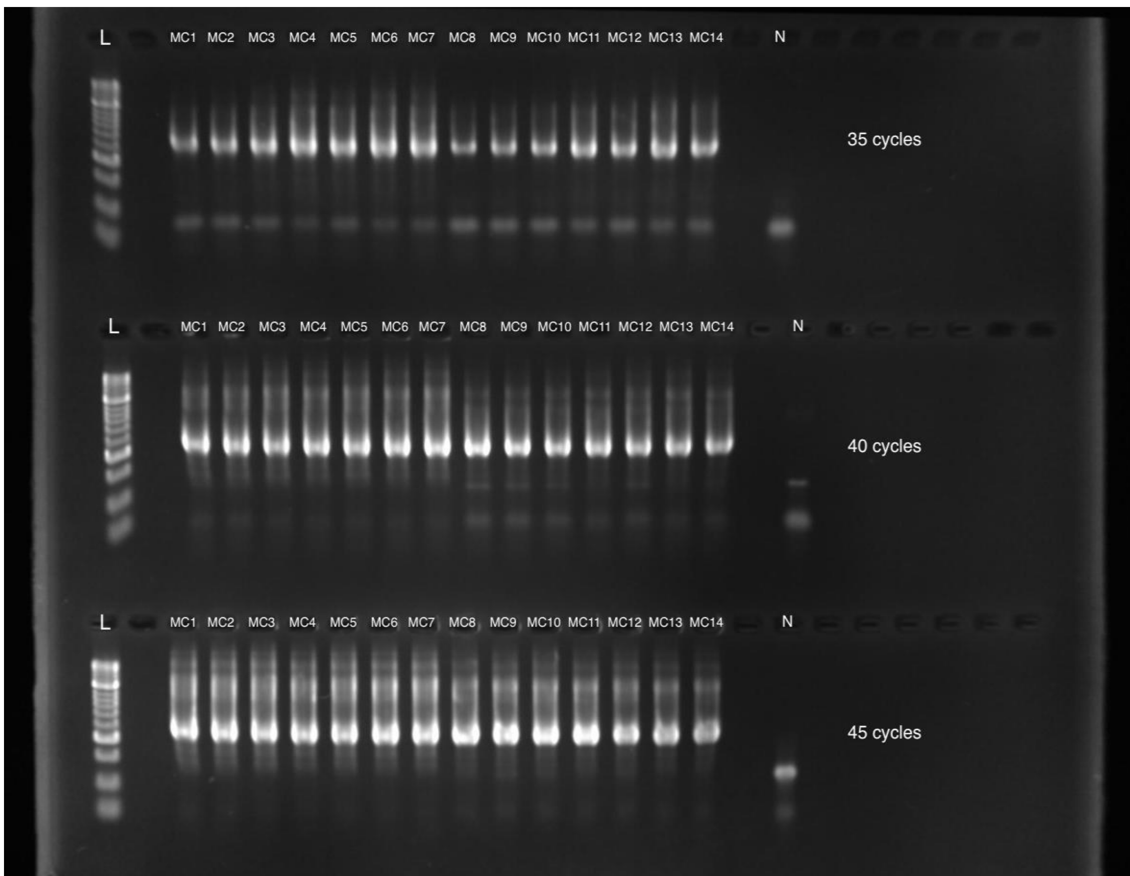


Figure S5.7: Gel images showing results of PCR cycle number optimisation for primers mICOIintF and jgHCO2198. PCR products were run on 2% agarose gels with Hyperladder™ 50bp (Bioline®, UK) molecular weight marker (L).

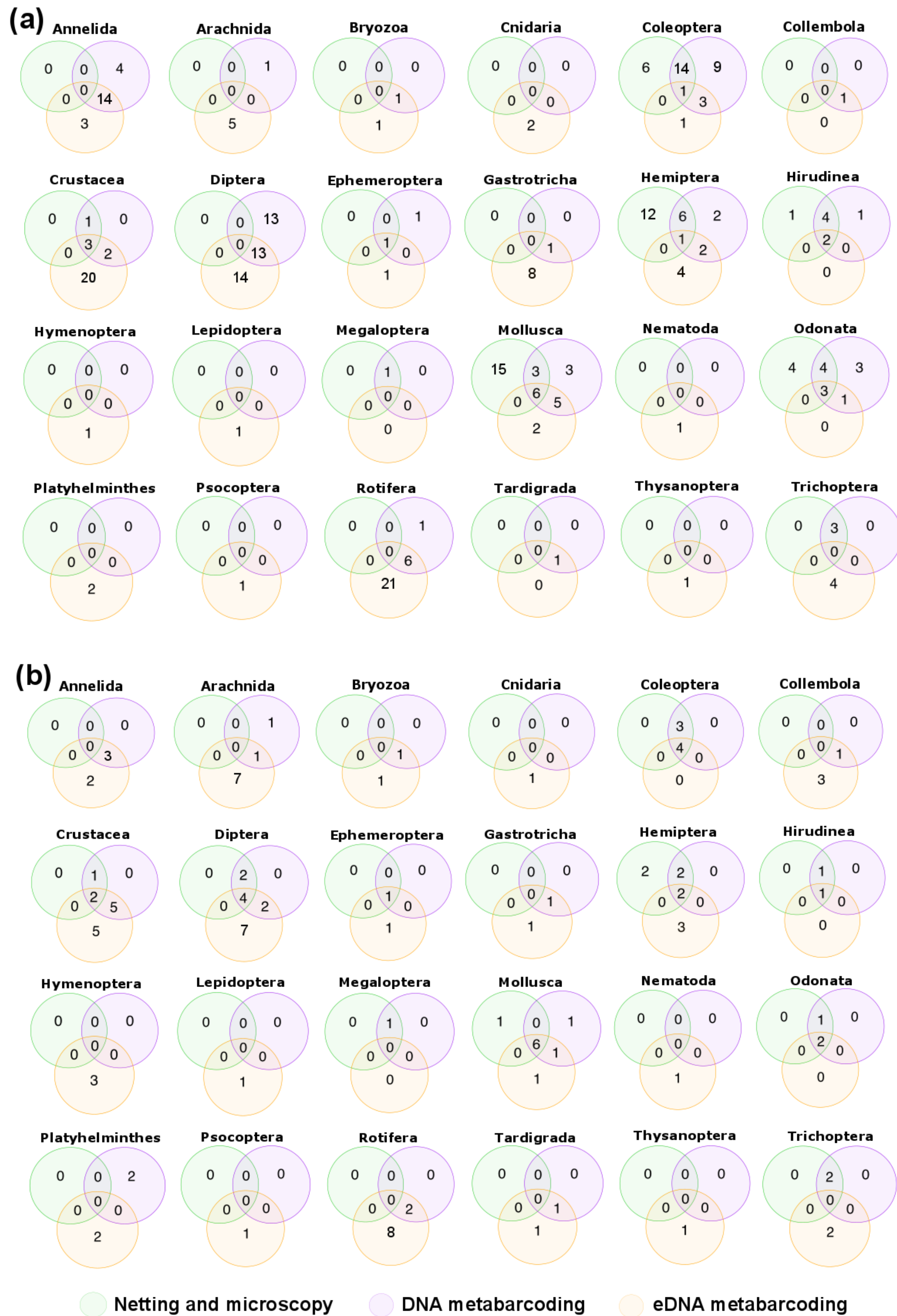


Figure S5.8: Venn diagrams which summarise the number of species (a) and families (b) detected within the major invertebrate groups across the 18 study ponds by each method of invertebrate assessment: sweep-netting and microscopy (green circle), DNA metabarcoding (purple circle), and eDNA metabarcoding (orange circle). Overlap in species or family detections between methods is displayed within circle intersections.

Appendix 5.5: Supplementary references

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bronner, I.F., Quail, M.A., Turner, D.J. & Swerdlow, H. (2009) Improved Protocols for Illumina Sequencing. *Current Protocols in Human Genetics*, **18**, 18.2.1–18.2.42.

Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Elbrecht, V. & Leese, F. (2017) Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science*, **5**.

Ficetola, G.F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P. & Pompanon, F. (2010) An *In silico* approach for the evaluation of DNA barcodes. *BMC Genomics*, **11**, 434.

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

Geller, J., Meyer, C., Parker, M. & Hawk, H. (2013) Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, **13**, 851–861.

Illumina, (2011) *Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System*. Illumina technical note.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonier, M., Kyle, F.J.K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S. & Willing, C. (2016) Jupyter Notebooks – a publishing format for reproducible computational workflows. In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. IOS Press, pp. 87–90.

Kitson, J.J.N., Hahn, C., Sands, R.J., Straw, N.A., Evans, D.M. & Lunt, D.H. (2019) Detecting host-parasitoid interactions in an invasive Lepidopteran using nested tagging DNA-metabarcoding. *Molecular Ecology*, **28**, 471–483.

Kozlov, A.M., Zhang, J., Yilmaz, P., Glöckner, F.O. & Stamatakis, A. (2016) Phylogeny-aware identification and correction of taxonomically mislabeled sequences. *Nucleic Acids Research*, **44**, 5022–5033.

Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T. & 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**, 34.

Magoč, T. & Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.

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.

Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, **4**, e2584.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Szitenberg, A., John, M., Blaxter, M.L. & Lunt, D.H. (2015) ReproPhylo: An Environment for Reproducible Phylogenomics. *PLoS Computational Biology*, **11**, e1004447.

Vamos, E.E., Elbrecht, V. & Leese, F. (2017) Short COI markers for freshwater macroinvertebrate metabarcoding. *Metabarcoding and Metagenomics*, **1**, e14625.

Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, **7**, 203–214.

Appendix 6

Appendix 6.1: Supplementary methods

eDNA metabarcoding workflow

A two-step PCR protocol was performed on eDNA samples at the University of Hull. Dedicated rooms were available for pre-PCR and post-PCR processes. Pre-PCR processes were performed in a dedicated eDNA laboratory, with separate rooms for filtration, DNA extraction and PCR preparation of sensitive environmental samples. PCR reactions were set up in an ultraviolet and bleach sterilised laminar flow hood. To minimise cross-contamination risk between samples, eight-strip PCR tubes with individually attached lids were used instead of 96-well plates (Port *et al.* 2016) and PCR reactions were sealed with mineral oil (Sigma-Aldrich Company Ltd, Dorset, UK) droplets (Harper *et al.* 2018a, b). PCR positive ($N = 2$) and negative controls ($N = 2$) were included on each PCR run to screen for sources of potential contamination. The DNA (0.05 ng/µL) used for the PCR positive control ($N = 16$) was *Maylandia zebra,* as this is an exotic cichlid not found in UK freshwater habitats. The negative controls ($N = 16$) substituted molecular grade sterile water (Fisher Scientific UK Ltd, UK) for template DNA.

During the first PCR, the target region was amplified using published 12S rRNA primers 12S-V5-F (5'-ACTGGGATTAGATACCCC-3') and 12S-V5-R (5'-TAGAACAGGCTCCTCTAG-3') (Riaz *et al.* 2011) that were validated *in silico* for all UK vertebrates by Harper *et al*. (2018a, b). Primers were modified to include MID tags, heterogeneity spacers, sequencing primers, and pre-adapters. During the first PCR, three replicates were performed for each sample to combat amplification bias. PCR reactions were performed in 25 µL volumes, consisting of 3 µL of template DNA, 1.5 µL of each 10 µM primer (Integrated DNA Technologies, Belgium), 12.5 µL of Q5® High-Fidelity 2x Master Mix (New England Biolabs® Inc., MA, USA) and 6.5 µL molecular grade water (Fisher Scientific UK Ltd, UK). PCR was performed on an Applied Biosystems® Veriti Thermal Cycler (Life Technologies, CA, USA) with the following thermocycling profile: 98 °C for 5 mins, 35 cycles of 98 °C for 10 s, 58 °C for 20 s and 72 °C for 30 s, followed by a final elongation step at 72 °C for 7 mins. PCR products were stored at 4 °C until replicates for each sample were pooled, and 2 µL of pooled PCR product was added to

0.5 µL of 5x DNA Loading Buffer Blue (Bioline®, UK). PCR product was visualised on 2% agarose gels (1.6 g Bioline® Agarose in 80 mL 1x Sodium borate) stained with ethidium bromide, and gels were imaged using Image Lab Software (Bio-Rad Laboratories Ltd, UK). A PCR product was deemed positive where there was an amplification band on the gel that was of the expected size (200-300 bp). PCR products were stored at -20 °C until they were pooled according to PCR plate to create sub-libraries for purification with Mag-BIND® RxnPure Plus magnetic beads (Omega Bio-tek Inc, GA, USA), following the double size selection protocol established by Bronner *et al*. (2009). Ratios of 0.9x and 0.15x magnetic beads to 100 µL of each sub-library were used. Eluted DNA (30 µL) was stored at -20 °C until the second PCR could be performed.

The second PCR bound pre-adapters, MID tags, and Illumina adapters to the purified sub-libraries. Two replicates were performed for each sub-library in 50 µL volumes, consisting of 6 µL of template DNA, 3 µL of each 10 µM primer (Integrated DNA Technologies, Belgium), 25 µL of Q5® High-Fidelity 2x Master Mix (New England Biolabs® Inc., MA, USA) and 13 µL molecular grade water (Fisher Scientific UK Ltd, UK). PCR was performed on an Applied Biosystems® Veriti Thermal Cycler (Life Technologies, CA, USA) with the following thermocycling profile: 95 °C for 3 mins, 8 cycles of 98 °C for 20 s and 72 °C for 1 min, followed by a final elongation step at 72 °C for 5 mins. PCR products were stored at 4 °C until duplicates for each sub-library were pooled, and 2 µL of pooled product was added to 0.5 µL of 5x DNA Loading Buffer Blue (Bioline®, UK). PCR products were visualised on 2% agarose gels (1.6 g Bioline® Agarose in 80 mL 1x Sodium borate) stained with ethidium bromide, and gels were imaged using Image Lab Software (Bio-Rad Laboratories Ltd, UK). Again, PCR products were deemed positive where there was an amplification band on the gel that was of the expected size (300-400 bp). Sub-libraries were stored at 4 °C until purification with Mag-BIND® RxnPure Plus magnetic beads (Omega Bio-tek Inc, GA, USA), following the double size selection protocol established by Bronner *et al*. (2009). Ratios of 0.7x and 0.15x magnetic beads to 50 µL of each sub-library were used. Eluted DNA (30 µL) was stored at 4 °C until normalisation and final purification.

Sub-libraries were quantified on a Qubit™ 3.0 fluorometer using a Qubit™ dsDNA HS Assay Kit (Invitrogen, UK) and pooled proportional to sample size and concentration. The pooled library was purified using the same ratios, volumes, and protocol as second PCR purification. Based on Qubit™ concentration, the library was diluted to 6 nM for quantification by real-time quantitative PCR (qPCR) using the

NEBNext® Library Quant Kit for Illumina® (New England Biolabs® Inc., MA, USA). The library was also checked using an Agilent 2200 TapeStation and High Sensitivity D1000 ScreenTape (Agilent Technologies, CA, USA) to verify secondary product had been removed successfully and a fragment of the expected size (330 bp) remained. The library was frozen and transported in a sterile portable freezer (-20 °C) to Centre for Ecology & Hydrology, Wallingford, for sequencing. The library was sequenced at 11.5 pM with 10% PhiX Control on an Illumina MiSeq® using 2 x 300 bp V3 chemistry (Illumina Inc., CA, USA).

Raw sequence reads were demultiplexed using a custom Python script then processed using metaBEAT (metaBarcoding and Environmental Analysis Tool) v0.97.11 (https://github.com/HullUni-bioinformatics/metaBEAT). Raw reads were quality trimmed from the read ends (minimum per base phred score Q30) and across sliding windows (window size 5bp; minimum average phred score Q30) using Trimmomatic v0.32 (Bolger, Lohse & Usadel 2014). Reads were cropped to a maximum length of 110 bp and reads shorter than 90 bp after quality trimming were discarded. The first 18 bp of remaining reads were also removed to ensure no locus primer remained. Sequence pairs were merged into single high quality reads using FLASH v1.2.11 (Magoč & Salzberg 2011), provided there was a minimum overlap of 10 bp and no more than 10% mismatch between pairs. Only forward reads were kept for pairs that could not be merged. A final length filter was applied to ensure sequences were reflected the expected fragment size (90-110 bp). Retained sequences were screened for chimeric sequences against a custom reference database for UK vertebrates (Harper *et al.* 2018a, b) using the uchime algorithm (Edgar *et al.* 2011), as implemented in vsearch v1.1 (Rognes *et al.* 2016). Redundant sequences were removed by clustering at 100% identity ('--cluster_fast' option) in vsearch v1.1 (Rognes *et al.* 2016). Clusters were considered sequencing error and omitted from further processing if they were represented by less than three sequences. Non-redundant sets of query sequences were then compared against the UK vertebrate reference database (Harper *et al.* 2018a, b) using BLAST (Zhang *et al.* 2000). Putative taxonomic identity was assigned using a lowest common ancestor (LCA) approach based on the top 10% BLAST matches for any query that matched a reference sequence across more than 80% of its length at minimum identity of 98%. Unassigned sequences were subjected to a separate BLAST search against the complete NCBI nucleotide (nt) database at 98% identity to determine the source via LCA as described above. The bioinformatic analysis has been deposited in the GitHub repository for reproducibility.

Data analysis

We tested the hypothesis that volume of water filtered or number of filters used may affect read counts. A hierarchical binomial Generalized Linear Mixed Model (GLMM) with the logit link function from the development version of the R package glmmTMB (Brooks *et al.* 2017), including volume and number of filters as fixed effects and species nested within wildlife park as a random effect, was used. Validation checks were performed to ensure all model assumptions were met where possible and absence of overdispersion (Zuur *et al.* 2009). Model fit was assessed visually and with the Hosmer and Lemeshow Goodness of Fit Test (Hosmer & Lemeshow 2000) using the R package ResourceSelection v0.3-2 (Lele, Keim & Solymos 2016). Predictions for each model were obtained using the *predict* function and upper and lower 95% CIs were calculated from the standard error of the predictions. Plots were produced using the R package ggplot2 v3.3.0 (Wickham, 2016).

Appendix 6.2: Supplementary results

Neither volume of water filtered (χ^2_1 = 2.141 ± 0.143) or number of filters used (χ^2_1 = 0.108 ± 0.742) had a significant effect on the proportional read counts based on the hierarchical model (θ = 0.221, χ^2_{76} = 16.798, P = 1.000, pseudo-R^2 = 20.57%). Proportional read counts somewhat decreased (-0.003 ± 0.002, Z = -1.389, P = 0.165) as water volume filtered increased (Fig. S6.4a), and marginally increased (0.232 ± 0.703, Z = 0.331, P = 0.741) where two filters were used for water filtration (Fig. S6.4b).

Appendix 6.3: Supplementary tables

Table S6.1: Behavioural observation data for species housed at wildlife parks, including date, time, weather conditions, behaviour, and frequency and duration of behaviour. Abbreviations for species are as follows: *Lutra lutra* (OTT), *Castor fiber* (BEAV), *Meles meles* (BAD), *Cervus elaphus* (DEER), *Lynx lynx* (LYNX), *Sciurus vulgaris* (SQ), and *Martes martes* (PM).

Location	Date	Start time	End time	Weather	Air temperature (°C)	Species	Enclosure	Behaviour	Frequency	Duration (hrs)
Wildwood Trust	18/09/2017	10:29	11:29	Cloudy	13	DEER	1	Drinking	9	9.37
								Feeding	3	22.37
								Defecating	1	0.5
								Sniffing	1	0.53
								Standing	6	30
								Walking	12	22.27
Wildwood Trust	19/09/2017	09:50	10:50	Sunny	17	DEER	1	Drinking	2	4.47
								Feeding	2	40.68
								Standing	1	34.8
								Walking	1	5.75
								Resting	1	6.7
Wildwood Trust	18/09/2017	11:38	12:38	Partial sun	16	LYNX	1	Scratching	1	0.08
								Urinating	2	3.42
								Standing	1	0.38
								Walking	9	18.12
								Walking	2	9.17
								Running	1	0.38
								Vocalising	4	4.58
								Resting	4	29.6
								Resting	1	1.32
								Grooming	4	5.03
								Grooming	1	0.08
								Not visible	1	1.88
								Not visible	2	48.25

Wildwood Trust	19/09/2017	09:30	10:30	Cloudy	12	LYNX	1	Drinking	3	1.13
								Urinating	1	0.01
								Not visible	1	10.52
								Resting	1	2.42
								Walking	3	46.8
								Walking	3	27.92
								Resting	1	0.77
								Grooming	1	0.45
								Other	1	0.97
								Not visible	2	30.95
Wildwood Trust	18/09/2017	11:36	12:32	Partial sun	16	PM	1	Immersed	10	0.35
								Drinking	2	0.03
								Urinating	1	0.03
								Other	19	0.67
Wildwood Trust	21/09/2017	09:26	09:56	Sunny	17	PM	1	Drinking	1	0.12
								Sniffing	7	6.83
								Standing	1	0.82
								Walking	10	8.45
								Resting	10	12.18
								Playing	9	5.22
								Not visible	1	0.98
Wildwood Trust	21/09/2017	09:56	10:26	Sunny	17	PM	2	Immersed	3	0.2
								Urinating	9	0.27
								Standing	1	0.23
								Running	4	29.57
								Resting	1	0.2
Wildwood Trust	18/09/2017	13:25	13:55	Partial sun	17	SQ	1	Feeding	5	3.77
								Walking	1	1.15
								Running	4	3.68
								Resting	2	0.8
								Drinking	2	0.93
								Running	4	23.83
								Resting	2	5.27

Wildwood Trust	18/09/2017	14:00	14:30	Partial sun	17	SQ	2	Drinking	1	1.4
								Feeding	4	15.5
								Running	3	13.1
								Drinking	2	0.57
								Feeding	3	1.07
								Running	6	27.37
Wildwood Trust	18/09/2017	13:45	14:45	Sunny	16	OTT	1	Swimming	12	14.25
								Standing	1	0.35
								Playing	2	5.28
Wildwood Trust	19/09/2017	09:21	10:21	Cloudy	12	OTT	1	Swimming	8	16.4
								Feeding	3	5.7
								Sniffing	5	8.33
								Not visible	4	35.43
Wildwood Trust	19/09/2017	19:54	05:04	Cloudy	9	BAD	1	Drinking	14	0.95
								Sniffing	36	8.43
								Walking	36	8.43
Wildwood Trust	19/09/2017	20:47	06:58	Clear	9	BEAV	1	Swimming	40	13.3
								Sniffing	1	0.33
								Standing	1	0.15
								Walking	12	1.52
Highland Wildlife Park	09/10/17	15:09	16:09	Partial sun	13	LYNX	1	Drinking	1	46
								Feeding	2	8.77
								Walking	6	19.82
								Resting	2	13.88
								Grooming	1	2
								Other	2	0.92
Highland Wildlife Park	NA	NA	NA	NA	NA	BEAV	1	Swimming	6	1.07
								Standing	2	0.5
								Sniffing	1	0.08
								Feeding	1	0.27

								Other	1	0.27
Highland Wildlife Park	09/10/17	09:55	12:25	Cloudy	12	DEER	1	Drinking	1	0.02
								Feeding	4	32.32
								Walking	7	16.62
								Resting	1	23
								Other	1	0.08
								Not visible	5	42.78

Table S6.2: Summary of directed, random, or other samples collected for each species at wildlife parks. Abbreviations for species are as follows: *Arvicola amphibius* (WV), *Lutra lutra* (OTT), *Castor fiber* (BEAV), *Erinaceus europaeus* (HH), *Meles meles* (BAD), *Cervus elaphus* (DEER), *Lynx lynx* (LYNX), *Sciurus vulgaris* (SQ), and *Martes martes* (PM).

Site	Species	Enclosure	Sample type	Number of samples	Volume filtered (mL)
Wildwood Trust	OTT	1	Targeted	5	500
			Passive	6	500
	WV	1	Other	1	250
		2	Other	1	250
	BEAV	1	Targeted	4	150
			Passive	5	150-200
		2	Passive	1	150
	HH	1	Other	1	250
		2	Other	1	250
	BAD	1	Targeted	3	500
			Passive	3	500
			Other	1	500
	DEER	1	Targeted	6	10-75
			Passive	4	25-150
	LYNX	1	Targeted	1	500
			Passive	2	500
			Other	1	500
	SQ	1	Other	1	250
		2	Other	1	250
		3	Other	1	250
		4	Other	1	250
	PM	1	Targeted	3	500
			Passive	2	500
			Other	1	500
		2	Targeted	1	500

			Passive	1	500
Highland Wildlife Park	SQ	NA	Other	1	500
	LYNX	1	Targeted	1	500
			Passive	2	500
			Other	1	500
	BEAV	1	Targeted	3	500
			Passive	3	500
			Other	1	100
	DEER	1	Targeted	5	50-200
			Passive	5	125-500

Table S6.3: Summary of exotic amphibian species housed in the wet laboratory at the University of Kent where water samples from Wildwood Trust were filtered, and number of sequence reads across eDNA samples assigned to these species.

Common name	Binomial name	Read counts
Golden mantella	*Mantella aurantiaca*	0
Mallorcan midwife toad	*Alytes muletensis*	0
Kaiser's spotted newt	*Neurergus kaiseri*	0
Mexican axolotl	*Ambystoma mexicanum*	0

Table S6.4: Summary of samples collected from natural ponds at locations where target species were confirmed as present.

Site	Date	Pond	Sample	Volume filtered (L)
Thorne Moors	17/04/2018	1	1	1.5
			2	1
			3	0.65
			4	0.8
			5	0.85
			6	0.8
			7	1
			8	1.5
			9	0.15
			10	0.1
		2	1	0.175
			2	0.4
			3	0.4
			4	0.5
			5	0.75
			6	0.75
			7	0.3
			8	0.3
			9	0.4
			10	0.6
Bamff Estate	20/04/2018	1	1	0.75
			2	0.45
			3	0.55
			4	1
			5	2

			6	0.9
			7	1
			8	0.95
			9	1.05
			10	0.6
		2	1	0.95
			2	0.95
			3	0.85
			4	1
			5	0.95
			6	0.95
			7	1.1
			8	0.95
			9	0.95
			10	0.95
Tophill Low Nature Reserve	23/04/2018	1	1	0.6
			2	0.625
			3	0.625
			4	0.675
			5	1.1
			6	0.85
			7	1
			8	0.625
			9	0.6
			10	0.65
		2	1	1
			2	1

		3	1
		4	1
		5	0.9
		6	0.9
		7	0.95
		8	0.95
		9	0.95
		10	0.85
24/04/2018	1	1	0.625
		2	0.8
		3	0.7
		4	0.65
		5	0.9
		6	0.8
		7	0.75
		8	0.65
		9	0.65
		10	0.625
	2	1	1
		2	0.9
		3	0.9
		4	1
		5	0.875
		6	0.825
		7	0.85
		8	1.1
		9	1.2
		10	1.1

25/04/2018	1	1	0.65
		2	0.75
		3	0.825
		4	0.55
		5	1
		6	0.8
		7	0.8
		8	0.65
		9	0.725
		10	0.55
	2	1	0.9
		2	0.9
		3	1
		4	0.925
		5	0.85
		6	0.775
		7	0.875
		8	1.1
		9	1.1
		10	1
26/04/2018	1	1	0.55
		2	0.775
		3	0.7
		4	0.7
		5	0.75
		6	0.85
		7	0.8
		8	0.65

		9	0.65
		10	0.725
	2	1	0.9
		2	0.9
		3	1
		4	0.9
		5	0.9
		6	0.775
		7	0.65
		8	0.75
		9	0.85
		10	0.7
27/04/2018	1	1	0.7
		2	0.65
		3	0.75
		4	0.65
		5	0.85
		6	0.8
		7	0.85
		8	0.7
		9	0.7
		10	0.75
	2	1	0.85
		2	0.8
		3	0.95
		4	0.75
		5	0.8
		6	0.85

	7	0.8
	8	0.85
	9	1.1
	10	0.9

Table S6.5: List of taxa detected in PCR positive controls by eDNA metabarcoding and corresponding taxon-specific false positive sequence threshold applied.

Taxonomic assignment	Common name	Threshold
Anas	Dabbling ducks	0.00067132
Anatidae	Ducks, geese, swans	0.000100995
Arvicola amphibius	European water vole	0.000342575
Aves	Birds	0.000054
Castor fiber	European beaver	0.003023912
Columba	Pigeons	0.0000877
Corvidae	Corvids	0.000081
Gasterosteidae	Sticklebacks	0.001862034
Homo sapiens	Human	0.000873784
Lynx lynx	Eurasian lynx	0.0000585
Martes martes	European pine marten	0.000906857
Mus musculus	Mouse	0.000107263
Passeriformes	Songbirds	0.0000202
Pelophylax ridibundus	Marsh frog	0.0000743
Phasianidae	Gamebirds	0.000107263
Phoxinus phoxinus	Common minnow	0.000092
Pungitius pungitius	Ninespine stickleback	0.026399055
Rana temporaria	Common frog	0.064393287
Sus scrofa domesticus	Domestic pig	0.000148423
Triturus cristatus	Great crested newt	0.001758274
unassigned	NA	0.009074043
Vanellus vanellus	Northern lapwing	0

Table S6.6: Summary of species detected using eDNA metabarcoding across all samples collected in this study.

Common name	Binomial name	Number of samples (N = 220)
Red-legged partridge	*Alectoris rufa*	2
Green-winged teal	*Anas carolinensis*	38
European eel	*Anguilla anguilla*	6
Grey heron	*Ardea cinerea*	14
European water vole	*Arvicola amphibius*	12
European bison	*Bison bonasus*	2
Cow	*Bos taurus*	44
Common toad	*Bufo bufo*	22
Common buzzard	*Buteo buteo*	3
Dog	*Canis lupus*	4
Roe deer	*Capreolus capreolus*	4
European beaver	*Castor fiber*	50
Red deer	*Cervus elaphus*	36
Atlantic herring	*Clupea harengus*	24
Rock dove	*Columba livia*	6
Stock dove	*Columba oenas*	6
Common quail	*Coturnix coturnix*	8
Grass carp	*Ctenopharyngodon idella*	1
Yellow-browed bunting	*Emberiza chrysophrys*	1
Horse	*Equus caballus*	28
European hedgehog	*Erinaceus europaeus*	1
European robin	*Erithacus rubecula*	8
Common moorhen	*Gallinula chloropus*	10
Eurasian jay	*Garrulus glandarius*	1
Three-spined stickleback	*Gasterosteus aculeatus*	18
Iceland gull	*Larus glaucoides*	3

Palmate newt	*Lissotriton helveticus*	4
Smooth newt	*Lissotriton vulgaris*	80
European otter	*Lutra lutra*	16
Eurasian lynx	*Lynx lynx*	22
European pine marten	*Martes martes*	16
Turkey	*Meleagris gallopavo*	3
European badger	*Meles meles*	25
Mouse	*Mus musculus*	11
Bank vole	*Myodes glareolus*	2
Eurasian water shrew	*Neomys fodiens*	7
Red-crested pochard	*Netta rufina*	18
European rabbit	*Oryctolagus cuniculus*	37
European smelt	*Osmerus eperlanus*	2
Sheep	*Ovis aries*	9
Great tit	*Parus major*	10
Marsh frog	*Pelophylax ridibundus*	11
Common pheasant	*Phasianus colchicus*	19
Common minnow	*Phoxinus phoxinus*	10
Eurasian magpie	*Pica pica*	7
Common pipistrelle	*Pipistrellus pipistrellus*	1
Ninespine stickleback	*Pungitius pungitius*	7
Common frog	*Rana temporaria*	20
Brown rat	*Rattus norvegicus*	10
Brown trout	*Salmo trutta*	12
Grey squirrel	*Sciurus carolinensis*	9
Red squirrel	*Sciurus vulgaris*	13
Common shrew	*Sorex araneus*	2
European sprat	*Sprattus sprattus*	5
Tawny owl	*Strix aluco*	1
Common starling	*Sturnus vulgaris*	2

Pig	*Sus scrofa domesticus*	45
European mole	*Talpa europaea*	1
Great crested newt	*Triturus cristatus*	100
Eurasian wren	*Troglodytes troglodytes*	1
Song thrush	*Turdus philomelos*	10
Northern lapwing	*Vanellus vanellus*	1

Table S6.7: Summary of detection rates for species which were detected by at least one survey method performed at six ponds across three sites in this study.

Species	Lifestyle	Field signs	Camera trapping	eDNA metabarcoding
Common pipistrelle (*Pipistrellus pipistrellus*)	Arboreal	0/6 (0%)	0/6 (0%)	1/6 (16.67%)
Grey squirrel (*Sciurus carolinensis)*	Arboreal	0/6 (0%)	0/6 (0%)	2/6 (33.33%)
Cow (*Bos taurus*)	Ground-dwelling	0/6 (0%)	0/6 (0%)	3/6 (50%)
Sheep (*Ovis aries*)	Ground-dwelling	0/6 (0%)	0/6 (0%)	4/6 (66.67%)
Pig (*Sus scrofa domesticus*)	Ground-dwelling	0/6 (0%)	0/6 (0%)	3/6 (50%)
Dog (*Canis lupus familiaris*)	Ground-dwelling	0/6 (0%)	0/6 (0%)	2/6 (33.33%)
Roe deer (*Capreolus capreolus*)	Ground-dwelling	4/6 (66.67%)	2/6 (33.33%)	3/6 (50%)
Red deer (*Cervus elaphus*)	Ground-dwelling	2/6 (33.33%)	1/6 (16.67%)	1/6 (16.67%)
Red fox (*Vulpes vulpes*)	Ground-dwelling	1/6 (16.67%)	3/6 (50%)	0/6 (0%)
Badger (*Meles meles*)	Ground-dwelling	1/6 (16.67%)	0/6 (0%)	0/6 (0%)
Bank vole (*Myodes glareolus*)	Ground-dwelling	0/6 (0%)	0/6 (0%)	1/6 (16.67%)
Common shrew (*Sorex araneus*)	Ground-dwelling	0/6 (0%)	0/6 (0%)	1/6 (16.67%)
Rabbit (*Oryctolagus cuniculus*)	Ground-dwelling	0/6 (0%)	0/6 (0%)	1/6 (16.67%)
Water vole (*Arvicola amphibius*)	Semi-aquatic	0/6 (0%)	0/6 (0%)	1/6 (16.67%)
Water shrew (*Neomys fodiens*)	Semi-aquatic	0/6 (0%)	0/6 (0%)	2/6 (33.33%)

Brown rat (*Rattus norvegicus*)	Semi-aquatic	0/6 (0%)	0/6 (0%)	1/6 (16.67%)
Beaver (*Castor fiber*)	Semi-aquatic	2/2 (100%)	2/2 (100%)	2/2 (100%)

Appendix 6.4: Supplementary figures

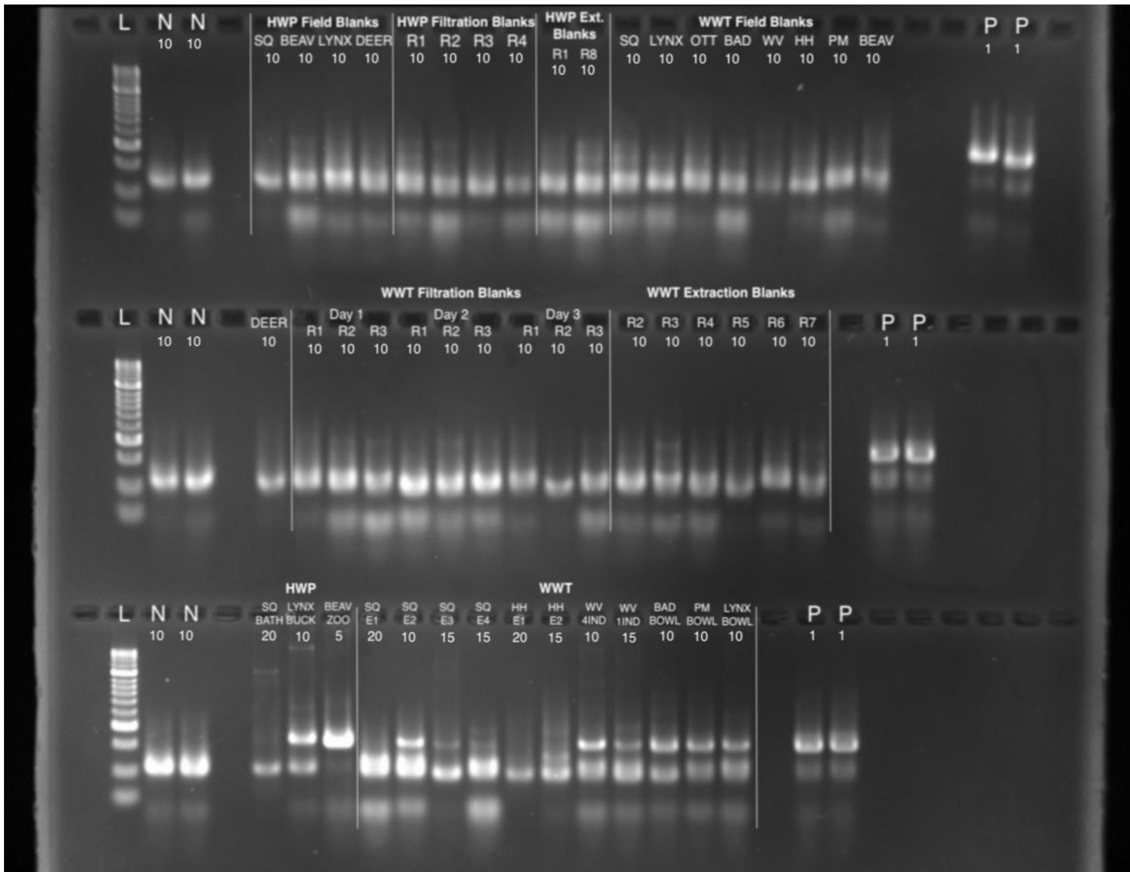


Figure S6.1: Example gel image of pooled first PCR products which were run on 2% agarose gels with Hyperladder™ 50bp (Bioline®, UK) molecular weight marker (L). PCR products were assigned an amplification score based on band strength (0 = no band, 1 = faint band, 2 = bright band, 3 = very bright band). These scores were used to determine how much product should be pooled to create each sub-library (0 = 20 µL, 1 = 15 µL, 2 = 10 µL, 3 = 5 µL). All blanks and PCR negative controls were pooled in consistent volumes (10 µL). Only 1 µL of each PCR positive control was pooled. Abbreviations for species are as follows: *Arvicola amphibius* (WV), *Lutra lutra* (OTT), *Castor fiber* (BEAV), *Erinaceus europaeus* (HH), *Meles meles* (BAD), *Cervus elaphus* (DEER), *Lynx lynx* (LYNX), *Sciurus vulgaris* (SQ), and *Martes martes* (PM).

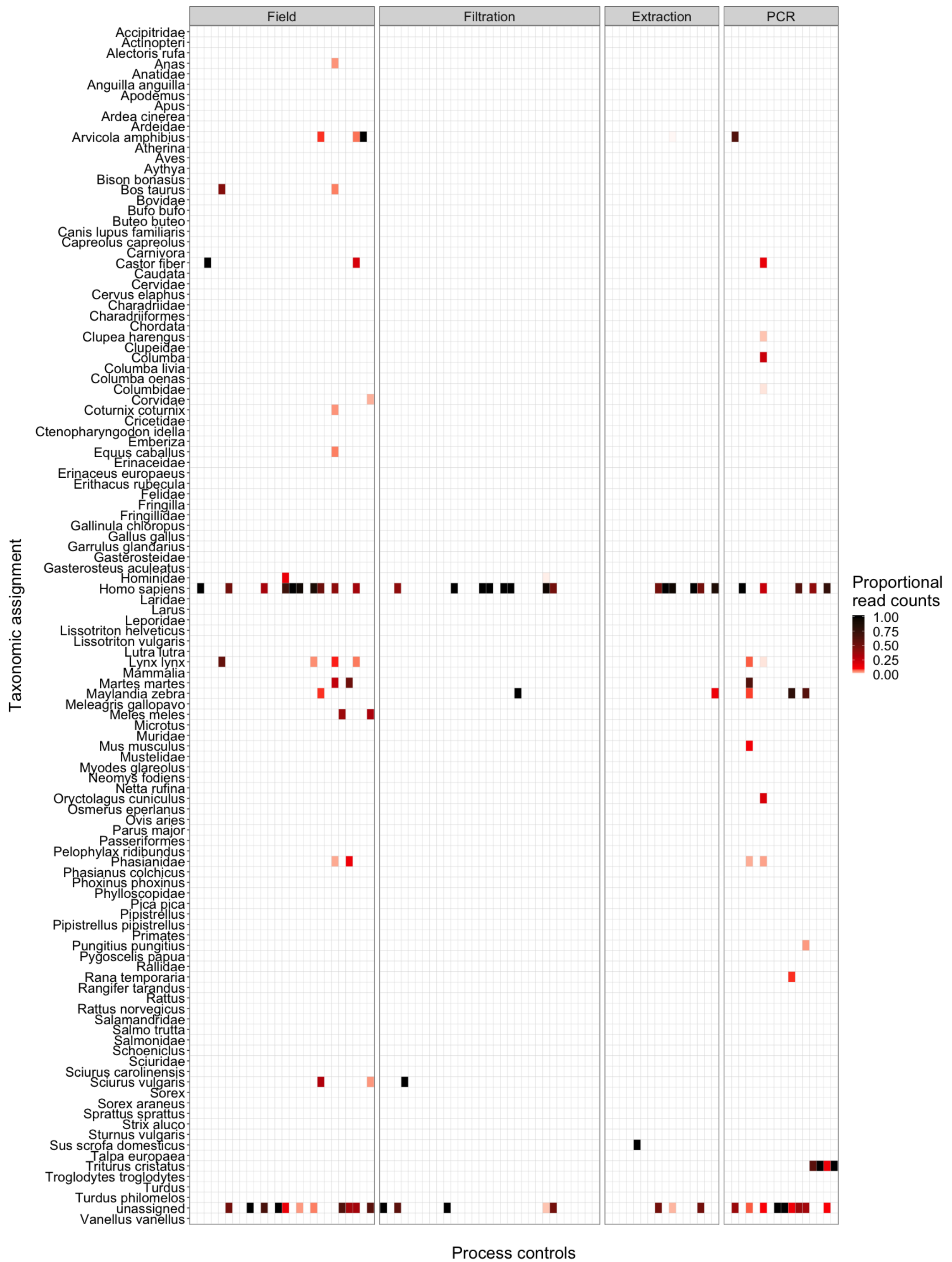


Figure S6.2: Heatmap showing the frequency of contamination in negative process controls (field blanks, filtration blanks, extraction blanks, and PCR negative controls). Assignments that were not detected in a given process control are coloured white.

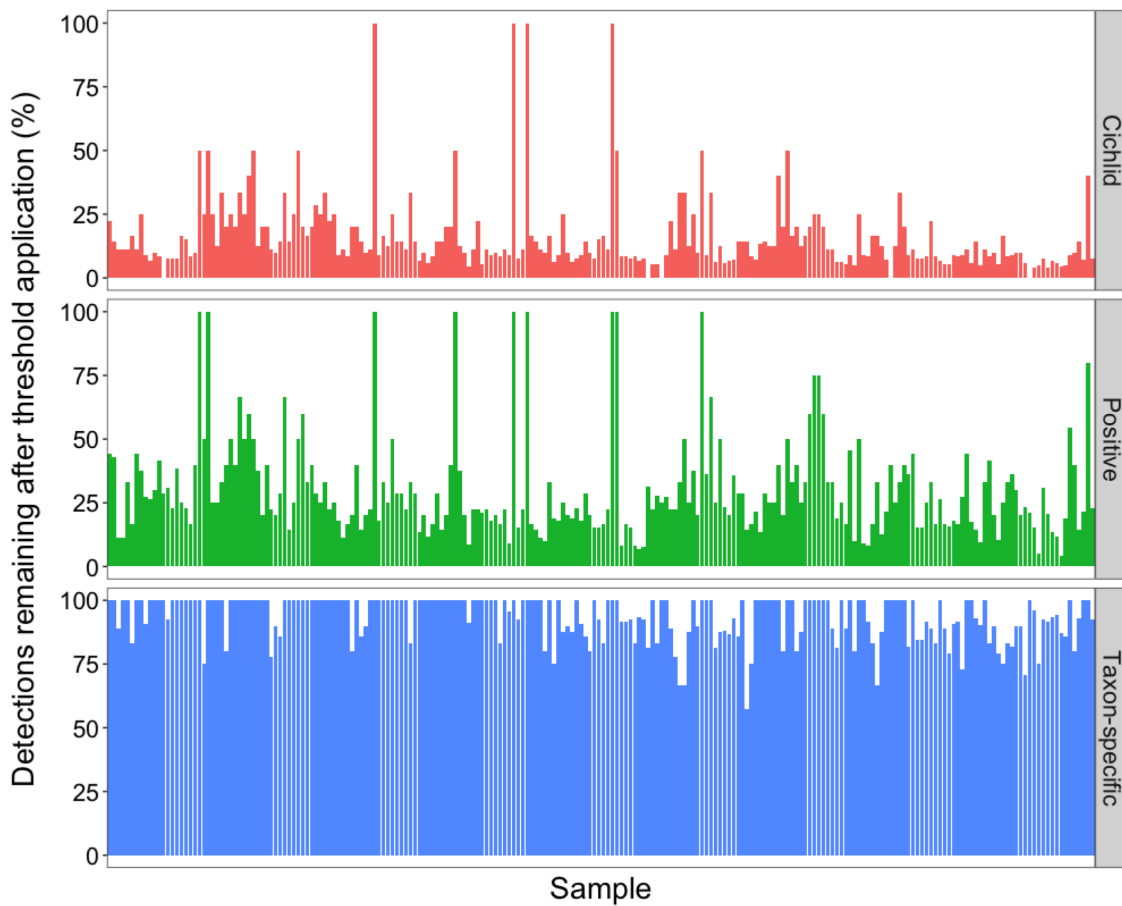


Figure S6.3: Barplot showing the impact of different false positive sequence thresholds on the proportion of taxa detected in each sample. The taxon-specific thresholds retained the most biological information, thus these were applied to the eDNA metabarcoding data for downstream analyses.

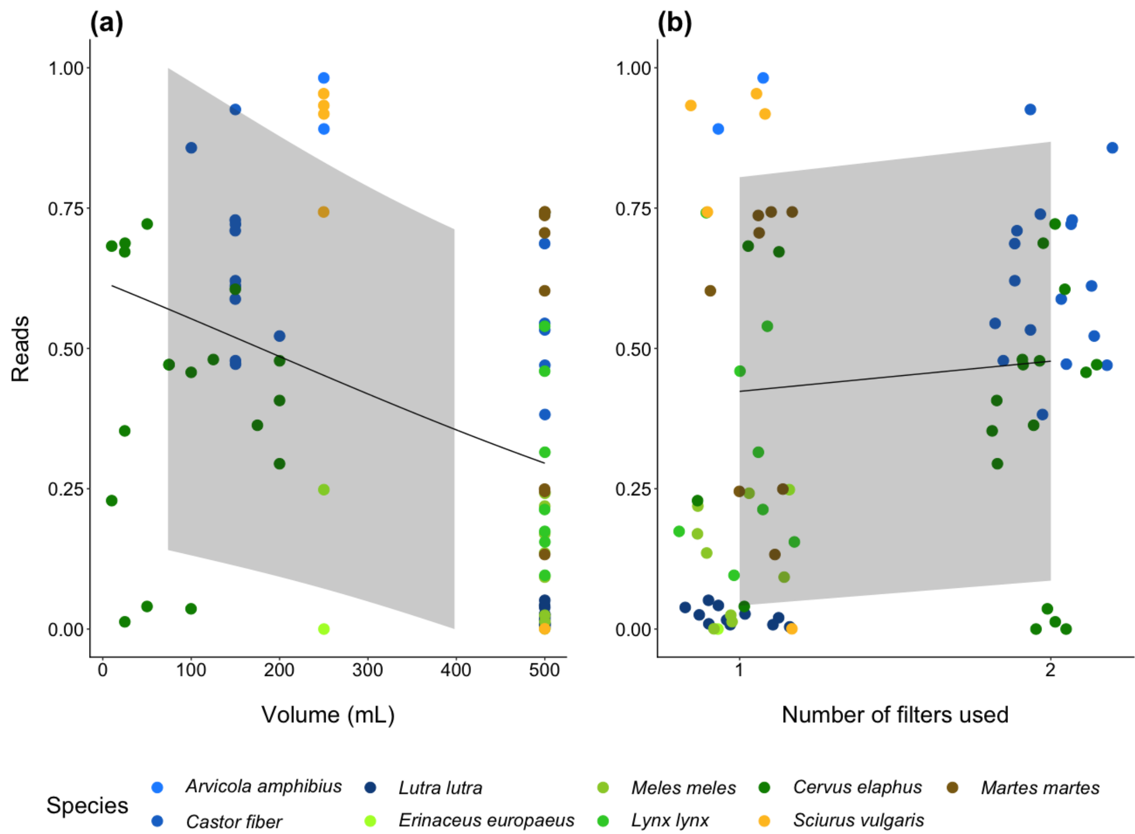


Figure S6.4: Relationship between the fixed effects (volume and number of filters) and response variable (proportional read count) as predicted by the binomial GLMM. The 95% CIs, as calculated using the predicted proportional read counts and standard error for these predictions, are given for each relationship. The observed data (points) are displayed against the predicted relationships (lines). Proportional read count marginally decreased as volume of water filtered increased **(a)**, but increased as number of filters used increased **(b)**.

Appendix 6.5: Supplementary references

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bronner, I.F., Quail, M.A., Turner, D.J. & Swerdlow, H. (2009) Improved Protocols for Illumina Sequencing. *Current Protocols in Human Genetics*, **18**, 18.2.1–18.2.42.

Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Machler, M. & Bolker, B.M. (2017) glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, **9**, 378–400.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Harper, L.R., Handley, L.L., Hahn, C., Boonham, N., Rees, H.C., Lewis, E., Adams, I.P., Brotherton, P., Phillips, S. & Hänfling, B. (2018a) Ground-truthing environmental DNA metabarcoding for ecological hypothesis testing at the pondscape. *bioRxiv*, 278309.

Harper, L.R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H.C., Gough, K.C., Lewis, E., Adams, I.P., Brotherton, P., Phillips, S. & Hänfling, B. (2018b) Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*). *Ecology and Evolution*, **8**, 6330–6441.

Hosmer, D.W. & Lemeshow, S. (2000) Multiple Logistic Regression. *Applied Logistic Regression*, pp. 31–46. John Wiley & Sons, Inc.

Lele, S.R., Keim, J.L. & Solymos, P. (2016) ResourceSelection: Resource Selection (Probability) Functions for Use-Availability Data. R package version 0.3-2.

Magoč, T. & Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.

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.

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

Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, **4**, e2584.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York, USA. http://ggplot2.org

Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, **7**, 203–214.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A. & Smith, G.M. (2009) *Mixed Effects Models and Extensions in Ecology with R*. Springer, New York, USA.