THE UNIVERSITY OF HULL

Environmental DNA: from detection of priority invasive species to monitoring entire macroinvertebrate communities in freshwater ecosystems

This thesis is submitted for the degree of Doctor of Philosophy

Rosetta Charlotte Blackman

November 2017

Candidates Declaration

I declare that the work submitted in this thesis is my own, except when otherwise stated. The work in this thesis involves collaborations with: Drew Constable (DC), Jessica Durkota (JD), Tim Goodall (TG), Hyun S Gweon (HG), Christoph Hahn (CH), Sean Ling (SL), Tim Jones (TJ), Daniel Read (DR), Peter Shum (PS) and Kerry Walsh (KW) and field work assistance from Frances Attwood (FA), Marco Benucci (MB), Charlotte Davy (CD), Lynsey Harper (LH), Helen Kimbell (HK), Andrew Sheard (AS) and Hayley Waton (HW)

Lori Lawson Handley (LLH) and Bernd Hänfling (BH) were involved with all work.

Chapter 2: MB, LH, HK, AS and HW assisted with field work

Chapter 3: AS and LH assisted with field work, PS advised on qPCR methods and development, SL assisted with laboratory work

Chapter 4: DC assisted with morphological identification, JD supplied extra samples for morphology and sequence identification and CH carried out phylogenetic analysis and assisted with bioinformatics

Chapter 5: FA, CD, AS assisted with field work, TG assisted with laboratory work, DR and HG carried out bioinformatics, TJ assisted with RIVPAC classifications.

I further declare that no part of this work has been submitted as part of any other degree. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Rosetta C Blackman, November 2017

Acknowledgments

Thanks firstly goes to my funders: The University of Hull, Environment Agency and the Gilson Le Cren Award from the Freshwater Biological Association.

To my supervisors, thank you for your support, ideas and advice. Lori, your knowledge, enthusiasm and dedication to this project has been a brilliant support and I'm certain without your belief in me I wouldn't be off on my next big adventure, thank you. Kerry, your encouragement, determination and persistence has made this work possible, I honestly would never have got here without you. Thank you also to my advisory panel; Helen Roy, Dan Read and Alice Hiley for all your encouragement and guidance and to everyone else who has given advice, collected samples with me (rain and shine) and been there through all of it – here's looking at you Tim and Fran, thanks also to Drew Constable for taking those first official *Gammarus fossarum* samples!

A special thank you to my lab mates and team "eDNA" over the last 3 years at Hull: Dan, Helen, Christoph, Amir, David, Jo, Lynsey, Graham, Marco, Cristina, Harriet, Peter, Hayley, James, Rob and last but by no means least Dr Paul Nichols of Canada. I am so grateful to have learnt so much from you all and I will look back on our time together and think of the many, many …cakes we have shared.

Thank you also to my friends and family who have supported and encouraged me in everything I've done – even the really silly things! A special thanks to my parents for being the perfectly wonderful people you are. And lastly the biggest thank you to Andrew, your support, kind heart and brilliant sense of humour has made this all possible. Special thanks, for keeping all those delirious, sleep deprived ramblings of mine quiet – there literally are "too many words", but I found them in the end.

Here's to our next adventure – now let's go buy that turkey.

Abstract

Freshwater ecosystems are among the most threatened habitats on Earth, facing challenges from a range of anthropogenic pressures. Accurate biodiversity assessment is essential to identify these pressures prior to irreversible damage. Current monitoring techniques for freshwater systems rely heavily on capture methods to infer the status of an ecosystem. However, these methods are often inefficient at detecting and identifying all species at a site and often miss those in low abundance. Emerging molecular methods such as environmental DNA (eDNA) could be a "game changer" for freshwater biodiversity monitoring. This thesis focuses on the application of eDNA for detection of invasive non-native species (INNS) and whole macroinvertebrate community assessment. Firstly, targeted eDNA PCR assays were developed for four priority freshwater INNS, and validated in mesocosm experiments and field trials. Targeted (PCR and qPCR) and passive (metabarcoding) eDNA approaches were then compared to traditional methods for detecting quagga mussels, *Dreissena rostriformis bugensis*. The targeted approaches were the most sensitive for detection of quagga mussels at low densities and both qPCR and metabarcoding showed correlations with mussel density. The power and utility of eDNA metabarcoding for detecting rare or unexpected taxa was then demonstrated by passive detection of a new INNS, *Gammarus fossarum*, in UK rivers. Finally, metabarcoding of both bulk DNA and eDNA from water and sediment was compared to the traditional method of macroinvertebrate sampling, to evaluate the potential of emerging molecular methods for ecological assessment. The results show metabarcoding approaches are not suitable to retrofit or replace existing methods of assessment, but provide an exciting opportunity for greater taxonomic identification and have the ability to detect a combination of taxa across groups, some of which are not utilised in current ecological assessment methods. This work has demonstrated a huge potential for eDNA methods to be applied to INNS monitoring and further our ability to carry out complete biodiversity assessment of waterbodies.

Table of Contents

Candidates Declaration I

Acknowledgments II

Abstract III

List of Figures IX

List of Tables XI

Chapter 1 General Introduction 1

1.1 Biodiversity monitoring needs 2

1.2 Invasive non-native species (INNS) 3

1.3 Current freshwater monitoring in the UK 3

1.4 Could environmental DNA revolutionise biodiversity monitoring? 4

1.4.1 Targeted approach: detecting single species 6

1.4.2 Passive approach: describing whole communities 8

1.5 Factors influencing the detection of eDNA 10

1.5.1 The ecology, physiology and anatomy of the organisms producing eDNA 10

1.5.2 Environmental variables that influence eDNA persistence 11

1.6 Aims and Objectives of the Thesis 12

1.6.1 Study species: 13

1.7 Ethics 16

1.8 Thesis rationale: developing molecular methods for use as ecological monitoring tools 16

Chapter 2 Targeting the invaders – targeted detection of four priority freshwater invasive non-native species using environmental DNA 19

2.1 Abstract 20

2.2 Introduction 21

2.3 Methods 25

2.3.1 Specimen sampling and tissue DNA extraction 25

2.3.2 Species specific primer development 25
2.3.3 Mesocosm experiments 26
2.3.4 Field trials 28
2.3.5 Data Analysis 30
2.4 Results 30
2.4.1 Primer specificity 30
2.4.2 Mesocosm experiments 33
2.4.3 Field trials 35
2.5 Discussion 36
2.5.1 Mesocosm trials 37
2.5.2 Field application 39
2.6 Conclusion 41
2.7 Supplementary Information I 42
2.8 Supplementary Information II 58
2.9 Supplementary Information III 59
Chapter 3 What is the optimal approach for monitoring aquatic invasive non-native species? A comparison of established and environmental DNA methods for monitoring the quagga mussel 68
3.1 Abstract 69
3.2 Introduction 70
3.3 Methods 73
3.3.1 Sample collection, DNA capture and extraction 73
3.3.2 Targeted Assays: standard PCR 74
3.3.3 Targeted Assays: qPCR assay development and validation in mesocosms 74
3.3.4 Passive detection: eDNA metabarcoding 76
3.3.5 Data analysis 78
3.4 Results 79

3.4.1 qPCR assay development and validation in mesocosms 79
3.4.2 Detection of *D. r. bugensis* by traditional and molecular approaches in the field . 81
3.5 Discussion 85
3.5.1 qPCR validation 85
3.5.2 Comparison of eDNA and kick-sampling for quagga mussel detection 86
3.5.3 Estimating quagga mussel abundance with eDNA 88
3.6 Conclusion 89
3.7 Supplementary Information I 90
3.8 Supplementary Information II – MIQE Checklist 98
Chapter 4 Detection of a new non-native freshwater species by DNA metabarcoding of environmental samples - first record of *Gammarus fossarum* in the UK 106
4.1 Abstract 107
4.2 Introduction 108
4.3 Methods 109
4.3.1 Metabarcoding surveys 109
4.3.2 Macroinvertebrate community sample processing 112
4.3.3 Environmental DNA sample processing 112
4.3.4 PCR, library preparation and sequencing 113
4.3.5 Specimen confirmation - microscopy and standard DNA barcode sequencing 114
4.3.6 Bioinformatics 115
4.3.7 Phylogeny 117
4.3.8 Comparison of data from eDNA/DNA and microscopy analysis 117
4.4 Results 118
4.4.1 Metabarcoding survey 118
4.4.2 Verification of Gammarus fossarum by microscopy 119
4.4.3 Verification of Gammarus fossarum by DNA barcoding 122
4.4.4 Phylogeny 122

4.5 Discussion 124
4.6 Conclusion 126
4.7 Supplementary Information I 127
Chapter 5 Understanding the potential application of metabarcoding macroinvertebrate communities for ecological assessment in rivers: a cautionary tale 140
5.1 Abstract 141
5.2 Introduction 141
5.3 Methods 144
5.3.1 Sampling and sample processing 144
5.3.2 Macroinvertebrate microscopy analysis 147
5.3.3 PCR, library preparation and sequencing 147
5.3.4 Bioinformatics 148
5.3.5 COI database 149
5.3.6 Data Analysis 149
5.4 Results 150
5.4.1 Kick samples 150
5.4.2 Comparison of sample types 152
5.4.3 Water Framework Directive river site classification 156
5.5 Discussion 159
5.5.1 Comparison of microscopy and metabarcoding 159
5.5.2 Comparison of molecular methods 161
5.5.3 Towards ecological classification with metabarcoding: how can we move forward? 163
5.6 Conclusion 164
5.7 Supplementary Information I 165
5.8 Supplementary information II 171
Chapter 6 Discussion 175

6.1 Overview of main findings 176
6.2 Targeted or passive approaches – which road to take? 178
6.3 What is needed from an eDNA monitoring method? 181
6.4 Towards optimal protocols for ecological monitoring with eDNA 182
6.5 Conclusion 185
6.6 Supplementary Information I 187
References 201

List of Figures

Figure 1.1: Environmental DNA workflow……..……5

Figure 1.2 Study species……13

Figure 2.1: Site locations for field trials……29

Figure 2.2: Results of *in vitro* primer testing.……..32

Figure 2.3: Species detection in mesocosm experiments…….…...33

Figure 2.4: Species detection in field experiments…….……...36

Figure S2.1: Agarose gel images from *Dreissena rostriformis bugensis* mescosom sampling….60

Figure S2.2: Agarose gel images from *Dreissena polymorpha* mescosom sampling……61

Figure S2.3: Agarose gel images from *Dikergammarus villosus* mescosom sampling…….....62

Figure S2.4: Agarose gel images from *Dikergammarus haemobaphes* mescosom sampling…..63

Figure S2.5: Agarose gel images from *Dreissena rostriformis bugensis* field samples collected from the River Wraysbury……64

Figure S2.6: Agarose gel images from *Dreissena polymorpha* field samples collected from the River Welland……...……65

Figure S2.7A and S2.7B: Agarose gel images from *Dikergammarus villosus* field samples……..……..…..66

Figure S2.8: Agarose gel images from *Dikergammarus haemobaphes* field samples collected from the River Nene.…… ……...…67

Figure 3.1: Sampling sites on the Wraysbury River…….……76

Figure 3.2: *Dreissena rostriformis bugensis* qPCR assay validation – mesocosm……..……...80

Figure 3.3: *Dreissena rostriformis bugensis* DNA signal recorded at field sites. ……..……84

Figure S3.1: qPCR product visualised on an agarose gel to show amplification……...……..……94

Figure S3.2. Agarose gel images from *D. r. bugensis* field samples collected from River Wraysbury……..……..95

Figure S3.3: qPCR (SYBR green) amplification plot……..…….96

Figure S3.4: Other taxa detected from the metabarcoding analysis……..97

Figure 4.1. Distribution of Gammaridae species detected during this study……111

Figure 4.2. Picture of *Gammarus fossarum* found in the River Taff……...…120

Figure 4.3 Maximum likelihood phylogenetic tree for the COI gene……..……121

Figure S4.1: Correlation of the % *Gammarus* biomass in the sample, and the percentage *Gammarus* sequence reads……..138

Figure S4.2. Frequency distribution of body length of male and female *Gammarus fossarum* individuals collected……...……139

Figure 5.1. Sampling sites used for this study. ……...……146

Figure 5.2: Summary of read counts from each molecular sampling method……151

Figure 5.3: sOTUs to Class/Order level for all four sampling methods. ……..……154

Figure 5.4: Comparison of methods. A and B Non-multidimensional scaling (NMDS) for all four sample types for all taxa (A) and for target taxa only (B). C. Sample type Venn diagram comparison of target taxa detected. ……155

Figure 6.1: Proposed monitoring strategy using current eDNA methods……180

List of Tables

Table 2.1. Primer pairs designed for this study and used for the detection of 4 target INNS……..……...……31

Table S2.1. GenBank accession numbers used for designing species specific primers for INNS target taxa..……..….42

Table S2.2: Primers designed and tested for species specific amplification of four target INNS……..……..43

Table S2.3: Tissue samples used in species specific primer testing…….……..50

Table S2.4: Specimen biomass pre and post mesocosm experiment, *Dreissena rostriformis bugensis*……..….51

Table S2.5: Temperature measurements during mesocosm experiment, *Dreissena rostriformis bugensis*…….……..51

Table S2.6: Specimen biomass pre and post mesocosm experiment, *Dreissena polymorpha*……..……..…52

Table S2.7: Temperature measurements during mesocosm experiment, *Dreissena polymorpha*……..……....52

Table S2.8: Specimen biomass pre and post mesocosm experiment, *Dikerogammarus villosus*……..……....53

Table S2.9: Temperature measurements during mesocosm experiment, *Dikerogammarus villosus*……..……53

Table S2.10: Specimen biomass pre and post mesocosm experiment, *Dikerogammarus haemobaphes*…….……54

Table S2.11: Temperature measurements during mesocosm experiment, *Dikerogammarus haemobaphes*……54

Table S2.12: Species specific primer mismatches. Details of mismatch amplification with reference sequences taken from Genbank. ……..…..…55

Table S2.13*: D. r. bugensis* mesocosm GLM…….……..…56

Table S2.14: *D. polymorpha* mesocosm GLM…….…56

Table S2.15: *D. villosus* mesocosm GLM……57

Table S2.16: *D. haemobaphes* mesocosm GLM……..….57

Table 3.1 A - C: Comparison of methods for the detection of *D. r. bugensis* at three sites on the Wraysbury River. ……82

Table S3.1: Sample sites used in this study……...……91

Table S3.2: *D. r. bugensis* field qPCR DNA copies GLM……93

Table S3.3: *D. r. bugensis* field qPCR DNA copies GLM……93

Table S3.4: *D. r. bugensis* field metabarcoding reads GLM……94

Table S3.5: MIQE guidelines checklist for qPCR assay development and reporting……..……..…99

Table 4.1. Specimen identification and identification method for morphologically identified and DNA barcoded specimens……...……123

Table S4.1: Specimen identification, identification method and site information for metabarcoding samples……...……128

Table S4.2 Information on specimens from own and published studies that were used in the phylogenetic tree.…… ……...……133

Table 5.1: Summary of read count and sOTUs. ……..……152

Table 5.2: ANOSIM results from paired sample type analysis. ……156

Table 5.3: Water Framework Directive classification…….……157

Table S5.1: Site details……166

Table S5.2: UK Environment Agency macroinvertebrate mixed-taxon level analysis guidance. ……168

Table S5.3: In vitro single specimen tissue samples tested using the Leray primer pair prior to sequencing……...……170

Tables S5.4 Venn diagram output from R, details of sOTUs in each sample type only and occurring in other sample types……...……172

Table S6.1 Questions and answers from the eDNA INNS consultation carried out in March 2017 ……188

Chapter 1 General Introduction

1.1 Biodiversity monitoring needs

Freshwater ecosystems are the most threatened habitats on earth, facing challenges from a range of anthropogenic pressures such as habitat loss, pollution, abstraction, invasive non-native species (INNS) and overexploitation (Strayer and Dudgeon, 2010, WWF, 2016). Accurate biodiversity assessment is essential to identify pressures on a system and carry out mitigation measures to prevent irreversible damage from occurring. Current monitoring techniques for freshwater systems rely heavily on sight or capture methods, such as trapping, netting or electrofishing, in order to identify species and infer the status of an ecosystem. However, these methods, and the subsequent identification of specimens, are often inefficient at detecting species in low densities and identifying small, juvenile and cryptic species (Lawson Handley, 2015). Molecular methods offer a promising complimentary approach for describing biodiversity which may have been previously missed or overlooked.

The International Barcode of Life (ibol.org) aimed at developing and standardising the use of DNA for species identification ("DNA barcoding") and has been established for some time (Valentini et al., 2009). More recently the use of DNA for species identification has been revolutionised by the analysis of DNA extracted from environmental samples (known as environmental DNA or "eDNA") (Taberlet et al., 2012a; Rees et al., 2014; Lawson Handley., 2015). The first eDNA studies focussed on describing historical fauna and flora from ancient sediments, where, given the right conditions, DNA can be preserved for tens of thousands of years (Willerslev et al., 2003). In aquatic systems however, eDNA samples are likely to be made up of contemporary biodiversity signals and be a combination of whole cells and degraded DNA fragments (Barnes et al., 2014). The persistence of eDNA within the system will however vary significantly, dependent on the abiotic and biotic properties of the environment (Rees et al., 2014; Barnes and Turner, 2016; Goldberg et al., 2015; Strickler et al., 2015) (discussed fully in section 1.5: Factors influencing the detection of eDNA). As this new method of ecological assessment allows the detection of species without prior collection (Ficetola et al., 2008) it makes the detection of species which are currently missed, due to low abundance, possible. This is of vital importance not only in the ecological assessment of communities but also the early detection of INNS which are often in low abundances prior to becoming established within a new system. The primary motivation of this thesis is to research the capabilities of eDNA to carry out ecological assessment and INNS detection. Here I discuss the relevance of the current thesis to freshwater biodiversity monitoring, with a focus on invasive non-native species and their wider

macroinvertebrate communities. I begin by evaluating the current status of freshwater monitoring, with a focus on the UK, then address whether eDNA can meet some of the current and future challenges, including factors that influence detection with eDNA. Finally, I introduce the study species targeted in this thesis and summarize the key aims of the 5 remaining thesis chapters.

1.2 Invasive non-native species (INNS)

Biodiversity in under considerable pressure, one of the five main drivers of global change and considered the greatest threat to biodiversity is the introduction and spread of INNS (Roy et al., 2014; WWF., 2016). Spread of INNS has been facilitated by increase in trade, tourism and transport in recent years (Hulme, 2009; Gallardo and Aldridge 2013a), and although often species which spread from their native range are not able to establish or colonise new areas (Manchester and Bullock, 2000), a small proportion of those that do, are likely to have a significant effect on native biodiversity, including increased competition for resources, predation on native species and vectors of new diseases (WWF, 2016). Once an INNS has become established attempts to remove them are often unsuccessful (Manchester and Bullock., 2000; Hulme., 2006; Howald et al; 2007). The most successful and cost-effective method to eradicating INNS is through prevention (screening pathways) and early detection (Hulme, 2006). However, much of the monitoring which takes place in freshwater e.g kick-sampling, is not aimed at detecting species in low abundance, (Dougherty et al., 2016) such as new INNS. Therefore, new INNS are only discovered once they reach a high enough density, which is often too late for eradication measures to be effective (Hulme, 2006). It is therefore vital that we explore new monitoring methods which will enable the detection of these species at an early stage of invasion when numbers are likely to be low and prior to establishing a self-sustaining population (Hulme, 2006).

1.3 Current freshwater monitoring in the UK

There are a number of local, national and international legislative requirements for monitoring freshwater macroinvertebrates. In Europe, the Water Framework Directive (WFD 2000/60/EC) is the overarching legislation that aims to improve and enhance the ecological status of waterbodies to 'good ecological' status in by 2027. For biological elements (macroinvertebrates, fish and phytobenthos) the term "good ecological status" is driven by Ecological Quality Ratios (EQRs), where observed community composition data are compared to the species which would be expected at a site if it was in pristine condition, taking into account river habitat (substrate), site location (altitude, distance from source) and water type (alkalinity). The waterbody is then

categorised as either bad, poor, moderate, good or high. By using the community composition of these elements, specific pressures can also be identified within a river, e.g. organic pollution, sediment impact and in some cases INNS. Although these monitoring requirements include recording any high impact INNS, there is very little direct INNS monitoring currently carried out. New or spreading INNS are often only detected by routine methods once a population has become established. With the addition of new European Union regulations "on the prevention and management of the introduction and spread of invasive alien species" (EU Regulation 1143/2014) the need for accurate forms of biodiversity assessment is not only desirable for the conservation of freshwater ecosystems but also a legislative requirement.

For macroinvertebrates, three sampling methods are approved under the WFD for classification of a water body. The method used is dependent on water depth and width: kick-net sampling (with one minute hand search), long-handled pond net sweep sampling and airlift sampling. However, these methods have their limitations. For example, the standard macroinvertebrate 3 minute kick net sample (with one minute hand search) method will only recover approximately 62% of families and 50% of species at a given site (Furse et al., 1981). Although increased sampling time does correlate with the number of families/species collected this is often not possible due to mounting cost and time constraints. Sample collection is also highly variable; with variation between samples, river habitat and sampler being noted as influences (Furse et al., 1981). These methods of physical collection are therefore likely to miss species in low abundance (Roy et al., 2014; Trebitz et al., 2017) and even if collected, morphological identification is highly dependent on the experience and skills of the analyst (Haase et al., 2006).

1.4 Could environmental DNA revolutionise biodiversity monitoring?

All organisms shed DNA into their environment, but the nature of the DNA, the absolute amount and rate at which it is shed varies greatly between species (Barnes and Turner, 2016). Our ability to detect eDNA depends on both the ecology and physiology of the species, as well as the environment in which they live (Barnes and Turner, 2016). Several studies have reviewed the use of eDNA as a tool for biodiversity assessment (Valentini et al., 2009; Darling and Mahon, 2011; Taberlet et al., 2012a; Taberlet et al., 2012b; Bohmann et al., 2014; Rees et al., 2014; Lawson Handley, 2015; Thomsen and Willerslev, 2015; Barnes and Turner, 2016; Creer et al., 2016 and Deiner et al., 2017). These reviews demonstrate the potential eDNA has for biodiversity monitoring including greater taxonomic resolution, detection of whole communities, and potential

cost and time benefits, and also highlight a variety of key areas which still need to be considered including: reaching a consensus on optimal methods both in the field and laboratory and for data analysis, better understanding the factors that influence eDNA detection in the field, and the need for collaboration between researchers and those who will apply these tools in the field. As the quantity of empirical studies grows a number of these elements are likely to be resolved, however a key area often overlooked is which approach may be best? This question forms a major focus of the current thesis, and below I introduce the two major approaches: a targeted approach for detection of single species, and a passive approach for describing whole communities (Fig. 1.1).

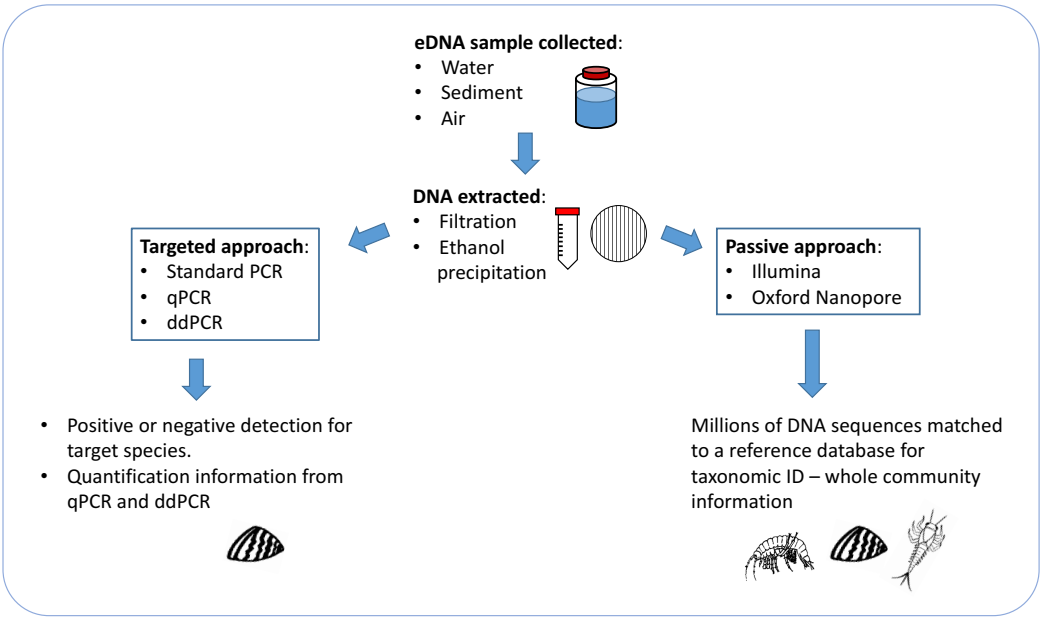


Figure 1.1: Environmental DNA workflow. Detailing the two downstream methods for eDNA analysis: a targeted or single species approach and passive or community based approach.

1.4.1 Targeted approach: detecting single species

The most commonly used method for species detection using eDNA is the targeted approach, which uses species specific primers during Polymerase Chain Reaction (PCR) to amplify the target DNA present in the sample (Goldberg et al., 2016). Over 100 species specific assays are now available (Lawson Handley, 2015; Hänfling et al., 2017). The first application of eDNA was to monitor the invasive American bullfrog, (*Rana catesbeiana = Lithobates catesbeianus*), using PCR in natural wetlands in France (Ficetola et al., 2008). This study showed detection at low densities of both tadpoles in mesocosms (as low as 0.3 individuals per litre) and adults in natural ponds, and, crucially, had no false positive detection. This study highlighted the future application of eDNA methods to detect invasive, rare and elusive species because of its sensitivity at low densities. Subsequently, a comparison demonstrated eDNA also had greater sensitivity to detect this species over traditional sampling (diurnal and nocturnal surveys). Bullfrogs were detected in 38 of the 49 ponds surveyed using eDNA compared to 7 of the 49 ponds using traditional methods (Dejean et al., 2012).

Standard PCR, as used in the American bullfrog study, relies on using an agarose gel to visualise the positive or negative detection in the form of the correct size band. Since this study a large number of PCR assays have been developed for INNS and other species of interest in both lentic and lotic systems. A number of other targeted studies have used quantitative real-time PCR (qPCR) which uses either dye-based detection or probe based technology to show amplification of target DNA in real time. For example, Takahara et al., (2012) showed a positive correlation between fish biomass and DNA copy number in both mesocosm and experimental ponds, and therefore the ability to provide quantitative estimates with eDNA. Finally, droplet digital (ddPCR) is a relatively new method based on microfluidics, in which samples are randomly partitioned into several thousand individual droplets which are then amplified by PCR, essentially allowing individual droplets of sample to be screened for target DNA (Nathan et al., 2014). ddPCR has only been trialled in a small number of eDNA studies, but with promising results (Nathan et al., 2014; Doi et al., 2015a; Doi et al., 2015b; Jerde et al., 2016, Simmons et al., 2015). For example, Simmons et al., (2015) used ddPCR to determine the presence of bighead carp, *Hypophthalmichthys nobilis*, but also showed detection of the target taxa was not possible with standard PCR. Both qPCR and ddPCR allow quantification of the target DNA within the sample and have greater sensitivity over PCR (Thomsen et al., 2012a; Nathan et al., 2014a), however PCR requires less specialist equipment, reagents and training than qPCR, does not require laborious construction of standard curves, and is generally cheaper. One study estimated that

reagents and consumables for standard PCR are 1.5-2 fold less than for qPCR, while PCR machines are 4 x cheaper than those for qPCR (Davison et al., 2016). All three methods rely heavily on the specificity of the primers, which involves extensive testing, development and validation via *in silico, in vitro* and *in situ* (Goldberg et al., 2016). Several studies show that primer specificity is key for successful detection and the prevention of false positives, with the risk of amplification of non-target species (Wilcox et al., 2013: Ardura et al., 2015a; Scriver et al., 2015; Goldberg et al., 2016). This is particularly important when trying to detect invasive or rare species which may be in low abundance and have closely related species within the same waterbody. Other issues such as the anatomy, physiology and ecology of the target species will influence the detection of target taxa, and in some cases, this can lower assay performance in the field (Mächler et al., 2014; Tréguier et al., 2014) (discussed fully in section 1.5.1: The ecology, physiology and anatomy of the organisms producing eDNA)

One drawback common to all targeted approaches is that development of species specific primers can be a lengthy process, requiring substantial investment in time and development, which must include stringent validation in both controlled and field conditions. A key consideration is whether to employ standard or quantitative approaches. Quantitative approaches are often favoured due to greater sensitivity, but in many cases the sensitivity of PCR may be adequate. Studies that have evaluated the detection rate of PCR have provided quite encouraging results. Jerde et al., (2011) for example, determined that silver carp, *Hypophthalmichthys molitrix*, could be detected by PCR from just seven copies of the target DNA per microlitre. Similarly, limits of detection were quite impressive for PCR assays (in the region of $4.6x10^{-4}$ ng/µl to $3x10^{-2}$ ng/µl) for four invasive fish species (*Lepomis gibbosus, Leucaspius delineatus, Pimephales promelas and Pseudorasbora parva*) (Davison et al., 2016).

Few studies have directly compared the performance of PCR and qPCR for species detection, however in a recent study on invasive zebra and quagga mussels (*Dreissena polymorpha* and *D. rostriformis bugensis*, respectively), PCR and qPCR detection methods were compared from laboratory, lake and river samples and PCR was found to be not only cheaper and simpler, but also more robust and less prone to false positives or false negatives than qPCR (De Ventura et al., 2017). The only study that has so far compared PCR, qPCR and ddPCR for species detection from environmental samples evaluated the method's ability to detect invasive round goby (*Neogobius melanostomus*) in controlled mesocosm experiments (Nathan et al., 2014). Water samples were collected from tanks with goby density varying from low to medium to high, over eight time

points ranging from 30 minutes to 24 hours after introduction of the fish (Nathan et al., 2014). Round goby DNA was detected in all samples using PCR, qPCR or ddPCR, demonstrating that the methods are comparable for presence-absence detection, at least under controlled conditions. More direct comparisons between methods are needed to determine whether this is more widely seen.

In summary, even PCR assays, which are less sensitive than qPCR or ddPCR, have been repeatedly shown to outperform established survey methods in terms of detection sensitivity (e.g. Ficetola et al., 2008; Dejean et al., 2012; Schmelzle et al., 2016). Although the most sensitive methods are often desirable for monitoring purposes, there is often a trade off between high sensitivity and false positives, and for some applications PCR, can be more suitable than qPCR or ddPCR (Stoeckle et al., 2016). use of a targeted approach is well advanced in terms of method development and validation but may be counterintuitive for the detection of new species which are not currently identified as high priority invasive or rare species. Similarly, this approach may not utilise the full potential of eDNA monitoring, which leads to the question: what further information can be gained from using a passive or whole-community approach?

1.4.2 Passive approach: describing whole communities

A passive monitoring approach utilises High Throughput Sequencing (HTS) technology to identify whole communities of taxa by generating millions of DNA sequences (Taberlet et al., 2012b). The most commonly used method is metabarcoding, which uses PCR to amplify target DNA, however in this case the primers target a group of taxa (i.e fish, invertebrates, mammals etc.) rather than a specific species. Metabarcoding libraries are run on an HTS platform such as an Illumina MiSeq or HiSeq, allowing a large number of samples to be simultaneously sequenced in a single run. This method produces large datasets of sequences which can then be compared to a reference database of sequences bioinformatically. Other technologies, such as PCR-free approaches and Oxford Nanopore's MinION platform are starting to be explored for eDNA work (Bai et al., 2014; Hänfling et al., 2017), but are still in their infancy. Here, I focus on metabarcoding, which was the method used in this thesis.

Several studies have demonstrated that a greater number of taxa can be detected with eDNA metabarcoding compared to traditional sampling (Thomsen et al., 2012a; Valentini et al., 2016; Hänfling et al., 2016; Port et al., 2016), and therefore metabarcoding could be informative as a

passive surveillance tool for INNS. Mahon and colleagues first demonstrated the effectiveness of metabarcoding for monitoring high risk pathways, in this case the bait trade, by collecting water samples from bait shops in the Great Lakes (Mahon et al., 2014; Nathan et al., 2015). Brown et al., (2016) also carried out an extensive metabarcoding survey of zooplankton from sixteen major Canadian ports, with the specific objective of INNS detection. Twenty-four non-native species were identified, of which eleven were previously undetected. One particular benefit of metabarcoding over a targeted approach is the additional information obtained for the whole community. Metabarcoding and other whole community approaches also allows studies to go much further than target species detection, for example in investigating impact of INNS on species richness or interactions. Furthermore, as demonstrated by Hänfling et al., (2016), eDNA metabarcoding data can be used to describe the ecological status of a water body, which could also lead to a greater understanding of community change and dynamics over time if an INNS becomes established in a new region (Simmons et al., 2015). However, of the very few studies which have compared target and passive approaches, both show metabarcoding to be less sensitive than targeted approaches (ddPCR and qPCR, Simmons et al 2015, Harper et al., 2017, respectively) it is therefore unclear if a passive approach is sufficiently sensitive to detect new and unexpected INNS in the field.

Metabarcoding is also appealing as the number of sequence reads generated could be considered a proxy for abundance, but both in principle and in reality, this relationship is complex. There are several opportunities for bias during the molecular and bioinformatics pipelines, such as primer bias and completeness of taxonomic reference databases. As with all molecular biodiversity monitoring, metabarcoding relies on primer specificity to all taxa within a group, however a major concern for this method is primer bias, i.e. amplifying certain taxa in preference to others, which has been highlighted by a number of studies (Deagle et al., 2014; Elbrecht and Leese, 2015), and will lead to species being missed. Secondly, downstream analysis of the data produced from metabarcoding requires bioinformatic analysis which can be highly complex and is heavily reliant on a reference database of DNA sequences to match the data produced to known species sequence records (Mahon et al., 2014; Simmons et al., 2015). Effort is being taken globally to add to existing reference databases such as the Barcode of Life (BOLD http://www.boldsystems.org/) and to link DNA barcode data with species record databases (e.g. as in the UK National Biodiversity Network (NBN) Atlas https://nbnatlas.org/), however obtaining funds for database construction remains a significant challenge for this work. In spite of these challenges, several studies have demonstrated a positive correlation between sequence reads and species biomass or rank abundance, although the relationship is not necessarily a simple linear one (Elbrecht and

Leese, 2015; Evans et al., 2016; Hänfling et al., 2016; Elbrecht et al., 2017; Stoeckle et al., 2017). A metabarcoding approach could be combined with site occupancy to provide an additional estimate of species abundance (Mackenzie et al., 2002; Mackenzie et al., 2003; Mackenzie et al., 2004). This combined approach was taken by Hänfling et al., (2016) in a study of the fish community of Lake Windermere, England. In this example, sequence read counts and site occupancy both correlated with rank abundance (assessed using long term data from gill-net and hydroacoustics surveys) but the latter correlation was stronger. This indicates that site occupancy could be a more appropriate surrogate for relative abundance than the number of sequence reads (Hänfling et al., 2016; Elbrecht et al., 2017; Stoeckle et al., 2017). A considerable amount of work is ongoing to try to improve abundance estimation both using PCR based and PCR-free approaches, but, at present, a combination of site occupancy information and read count seem to offer a reasonable approximation.

1.5 Factors influencing the detection of eDNA

Several studies have highlighted our need to gain a greater understand of how eDNA behaves within the system being sampled in order to understand the samples we are collecting. (Ficetola et al., 2008; Rees et al., 2014; Roussel et al., 2015; Barnes and Turner, 2016). eDNA comes from a range of sources including urine, faeces, sloughed cells, gametes and decaying material, and it is likely to be a combination of complete cells and degraded DNA (Turner et al., 2014; Barnes et al., 2015), all of which will break down quickly due to environmental factors (Turner et al., 2014). However, the primary factor influencing the detection of eDNA is the organism: its ecology, physiology and anatomy. Both factors are discussed in detail below.

1.5.1 The ecology, physiology and anatomy of the organisms producing eDNA

DNA production is likely to vary depending on the species being studied and as little is currently known about the rates of DNA shedding by individual species, this remains a relatively unexplored area of study. Several studies have focused on the detection of fish and amphibians (Jerde et al., 2011; Thomsen et al., 2012a; Pilliod et al., 2013; Tréguier et al., 2014; Klymus et al., 2015; Jane et al., 2015; Roussel et al., 2015;) as it is thought these groups readily shed DNA into their environment via sloughed cells and mucus and therefore may produce more eDNA than other taxa (Barnes et al., 2015). Although a number of macroinvertebrates have been successfully detected using both targeted and metabarcoding techniques (Thomsen et al., 2012a; Deiner et al., 2014; Deiner et al., 2016), comparatively low amounts of extracellular DNA in the water is a

major limitation for detection of invertebrates compared to fish and amphibians (e.g. Tréguier et al., 2014; Mächler et al., 2014). Both Mächler et al., (2014) and Tréguier et al., (2014) failed to amplify target DNA in field samples, despite detection with traditional methods, stating taxa physiology as a possible cause. For example, detecting the invasive red swamp crayfish, *Procambarus clarkii* was challenging when the species was in low abundances. Successful detection of crayfish has been linked to moulting behaviour, which is not constant and therefore will influence false negative results (Tréguier et al., 2014; Dunn et al., 2017). This has important implications when trying to detect freshwater invertebrate INNS, particularly at the early stages of invasion. Quantitative eDNA may reflect biomass or abundance which is particularly promising, however Klymus et al., (2014) showed eDNA shedding rates were also linked to behaviour and therefore results may depend strongly on seasonal changes in behaviour such as breeding seasons.

1.5.2 Environmental variables that influence eDNA persistence

In line with the differing rates of DNA production, eDNA is also subject to a range of environmental pressures which are likely to affect both the persistence and degradation of DNA. Detection success is also highly variable between systems, with persistence being a matter of days in marine systems (Thomsen et al., 2012a) and up to 3 weeks in freshwater ponds (Dejean et al., 2011), there are also further variables to consider (Rees et al., 2014; Goldberg et al., 2015; Strickler et al., 2015; Barnes and Turner, 2016;). Firstly, confounding factors such as high temperatures, UV and alkalinity have all been attributed to faster degradation (Strickler et al., 2015), although a more neutral pH and lower UV would stimulate microbial diversity which has also been to influence DNA persistence (Zhu 2006; Pilliod et al., 2013; Strickler et al., 2015). Although the influences of such environmental variables are likely to be linked e.g UV exposure and water temperature (Pilliod et al., 2013), full understanding of all the potential influences on eDNA as it moves through a lotic system is lacking and further research is still needed (Barnes and Turner, 2016).

In this thesis, I focus on eDNA within lotic systems, which feature variable flow and associated substrates. Identifying how target DNA is transported through these systems is essential for developing appropriate sampling strategies and methods (Klymus et al., 2015; Shogren et al., 2017). The availability of eDNA within lentic systems may reflect processes such as diffusion and circulation of water and so eDNA is readily available to samplers, whilst in lotic systems eDNA availability is likely to be influenced by hydrological dynamics (Jane et al., 2015; Jerde et

al., 2016; Shogren et al., 2017). These studies also highlight the influence of changing substrate type; as eDNA flows through the system it will be subject to increased retention in the substrate associated with slower flows, e.g finer substrate. However, by sampling rivers we could also see spate and high flows which will influence re-suspension of DNA and these conditions are likely to lead to greater dilution of eDNA although it is also likely to travel a greater distance, enhancing the opportunity to detect it (Jane et al., 2015).

Despite these influences target eDNA has been documented as persisting downstream in lotic systems for up to 9 km (Deiner and Altermatt., 2014), although a similar study of fish noted detection was limited to 293.5m (Jane et al., 2015), and although these studies further describe great variety in eDNA persistence, it is clear the eDNA collected in samples are likely to be a representation of the species at the site being sampled and also those present higher up the catchment. The effects of flow and river dynamics were eloquently summarised by Deiner et al., (2016) as acting as a "conveyor belt" for both aquatic and terrestrial biodiversity information and therefore, could offer a large-scale picture of community diversity within a catchment or identify the presence of a species of interest e.g. INNS. In a rapidly developing scientific field, comparisons of molecular methods with our existing monitoring tools are a key starting point in order to determine how eDNA reflects the ecological information we are currently collecting, and then how these new tools can be applied to routine monitoring. Furthermore, examination of both targeted and passive approaches is also needed in order to determine the most appropriate method for both community composition and the detection of target taxa e.g INNS.

1.6 Aims and Objectives of the Thesis

The aim of this thesis is to determine how eDNA can be used for monitoring biodiversity in lotic systems. A large driver of this thesis is the early detection of INNS and to develop methods which are appropriate to monitor of high risk pathways for INNS. Here I introduce the four priority species and then outline the specific aims of each chapter.

1.6.1 Study species:

In this thesis, I explore the use of eDNA for both community assessment and target taxa. The four INNS I will focus on are from the Ponto-Caspian region and Black Sea: quagga mussels, *Dreissena rostriformis bugensis* (Andrusov, 1987); zebra mussels, *Dreissena polymorpha* (Pallas, 1771); killer shrimp, *Dikerogammarus villosus* (Sowinsky, 1984) and demon shrimp, *Dikerogammarus haemobaphes* Eichwald, 1843) (Fig. 1.2). Both *D. r. bugensis* and *D. villosus* are high priority INNS, due to the current spread of *D. polymorpha* and *D. haemobaphes* these species are of lower priority to regulators but due to the morphological similarities of each species pair, biosecurity and prevention of further spread of all species remains an important aim by the UK Environment Agency. Below I briefly outline the life-history and ecology of the four species and why they are high priority INNS for the UK.

Figure 1.2 Study species: A - quagga mussels, *Dreissena rostriformis bugensis*; B - zebra mussels, *Dreissena polymorpha*; C - killer shrimp, *Dikerogammarus villosus*; D – demon shrimp – *Dikerogammarus haemobaphes*.

Dreissenidae mussels

Dreissena polymorpha is widespread and common in the UK and it's arrival in the 1820's can be directly linked to the timber trade and transport through Western Europe, via the Northern and Central corridor (Bij de Vaate et al., 2002; Quinn et al., 2014). *Dreissena rostriformis bugensis* however, is a much more recent invader, with the first UK record from 2014 in the River Wraysbury. Subsequent surveys showed *D. r. bugensis* was also present in neighbouring reservoirs in the area, having been transported by water transfers. Both mussels are described as "ecological engineers" (Karatayev et al., 2002; Karatayev et al., 2007; Roy et al., 2014). As opposed to the many other INNS, Dreissenid mussels are capable not only of negative impacts on all trophic levels within the ecosystem they invade, including increasing cyanobacteria blooms and decreasing numbers of native Unionidae numbers (Ricciardi et al., 1996; Karatayev et al., 2002), but these INNS also cause a significant impact to water companies who need to clear mussel shells from their infrastructure at significant financial cost (Connelly et al., 2007).

Dreissenid mussels are able to spread rapidly and colonize new waterbodies due their unique ecology (Timar and Phaneuf, 2009). Like other mussels, Dreissenids have a free-floating planktonic veliger life stage, during which young can be dispersed over a large area downstream (Ricciardi et al., 1995; Karatayev et al., 2002; Karatayev et al., 2015). However, Dreissenid mussels are also adapted to tolerate environmental extremes and survive out of water for up to 15 days (Ricciardi et al., 1995; Anderson et al., 2015). In part, this is aided by their ability to produce protein-based byssal strands which secure their shell to hard surfaces, which can be a significant aid to transportation and establishment (Ricciardi et al., 1998; Karatayev et al., 2002; Aldridge et al., 2004; Timar and Phaneuf, 2009; Peyer et al., 2009).

Monitoring and preventing the spread of *D. r. bugensis* is now a priority within the UK as it is likely to be able to invade a wider range of habitats than the *D. polymorpha*, due to its wider tolerance of environmental variables (Nalepa et al., 2010; Quinn et al., 2014). *Dreissena rostriformis bugensis* are also able to spawn at lower temperatures than *Dreissena polymorpha* (Roe and MacIsaac, 2011), which suggests they will potentially thrive in the cool UK climate. Therefore, this relatively recent arrival now poses a new challenge to UK regulators due to the morphological similarities of the two species. It is hoped that regulators will be able to use new

eDNA tools to monitor high risk pathways, such as water and fish transfers and detect any potential invasions prior to establishment.

Dikerogammarus species

Dikerogammarus villosus and *D. haemobaphes* have spread in a similar way to the Dreissenid mussels, arriving in Germany by the late 1990s. The first UK record of *D. villosus* was from Grafham water, in September 2010, where the species was already in high densities in the reservoir when identification was confirmed (MacNeil et al., 2010a). Strict biodiversity controls were put in place (Check, Clean, Dry, www.nonnativespecies.org) and subsequent spread has been limited to five locations in the UK. *Dikerogammarus haemobaphes* however, has rapidly colonized British waterways since its discovery in September 2012, (Environment Agency, 2017) spreading successfully through the river and canal networks.

Dikerogammarus villosus is noted for its exceptional predatory capabilities (Dick et al., 2002; MacNeil et al., 2010a), its high reproductive output (MacNeil et al., 2010a) and also known to survive for up to six days out of water, allowing for extensive transportation on kit and in ballast water (Martens & Grabow, 2008). *D. haemobaphes*, although less well studied than *D. villosus*, has a similar trophic level to *D. villosus* but with less predatory impacts (Bacela-Spychalska and Can Der Velde, 2013; Bovy et al., 2015). Both species are documented as having significant negative effects on the macroinvertebrate community, including outcompeting native Gammaridae for resources (Dick et al., 2002; MacNeil et al., 2010a, MacNeil et al., 2012).

Preventing the further spread of *D. villosus* in the UK is a key biosecurity priority and an area which may benefit from eDNA methods. Although *D. haemobaphes* has spread successfully through canal and river networks, effective monitoring is critical to prevent further established populations. Like the Dreissenid mussels, the *Dikerogammarus* species are morphologically very similar and although the species seem to exhibit similar traits as INNS, our knowledge of *D. haemobaphes* is limited and therefore early detection and accurate speciation is a high priority for regulators.

1.7 Ethics

Specimens of all four study species were collected at the following sites: *D. r. bugensis*, Wraysbury River, UK Grid. Ref. TQ 02680 73204; *D. polymorpha* and *D. haemobaphes*, Rutland Water, SK 92956 05963; *D. villosus*, Grafham Water, TL 15081 67289. Experiments were approved by the ethical review committees of the School of Environmental Sciences and the Faculty of Science and Engineering at the University of Hull (reference number U086).

1.8 Thesis rationale: developing molecular methods for use as ecological monitoring tools

As discussed previously the potential uses of molecular methods are staggering, but it is only with methodical testing, comparison and exploration we can begin to understand how these methods can be used for community assessment and targeted species detection. The primary aim of this thesis is to determine to what extent molecular methods, both passive and targeted approaches can contribute to both whole biodiversity assessment of macroinvertebrates and the detection of INNS in lotic waterbodies in the UK. Both methods were used to explore these areas by firstly developing species specific assays for the four priority INNS and then comparing molecular and traditional methods both for target species detection and whole macroinvertebrate community assessment. The thesis is divided into six chapters:

Chapter 2: Targeting the invaders – targeted detection of four priority freshwater invasive non-native species using environmental DNA

The aim of this chapter was to design, develop and validate species specific primers for four high priority INNS in the UK. Assays were developed for all four species and tested in a series of mesocosm experiments with three different density treatments. This was followed with field experiments to validate eDNA assays as a suitable monitoring tools in the field. I compared the field samples with traditional monitoring techniques to demonstrate the increased sensitivity and greater effect using eDNA has on the detection of INNS. This chapter has been submitted to PeerJ.

Chapter 3: What is the optimal approach for monitoring aquatic invasive non-native species? A comparison of established and environmental DNA methods for the monitoring quagga mussel

In this chapter I compare both traditional sampling methods with molecular tools, both passive and targeted, for the detection of quagga mussel, *Dreissena rostriformis bugensis*. To continue the development of the targeted approach for *D. r. bugensis*, I extended the targeted approach developed in Chapter 2 for dye-based qPCR then validated the primer set in a mesocosm experiment, to evaluate the rate of DNA copy number production and degradation. Standard PCR, qPCR and metabarcoding were then compared to traditional sampling in the field along a stretch of the River Wraysbury to evaluate the sensitivity of the four approaches and the possibility of using qPCR and metabarcoding for abundance estimation.

Chapter 4: Detection of a new non-native freshwater species by DNA metabarcoding of environmental samples - first record of *Gammarus fossarum* in the UK

This chapter details the discovery of a new non-native species, *Gammarus fossarum* via metabarcoding. This species was identified during the initial data analysis of chapter 5 and lead to a number of new surveys in order to collect evidence, both morphologically and for DNA barcoding, to validate this new find. This chapter demonstrates the important use of passive approaches as opposed to targeted approaches for the detection of new INNS. This chapter was published in Aquatic Invasions in 2017 (Blackman et al., 2017a).

Chapter 5: Understanding the potential application of metabarcoding macroinvertebrate communities for ecological assessment in rivers: a cautionary tale

In the final data chapter, I compare traditional methods (kick-sampling followed by morphotaxonomy) with metabarcoding of three sample types (bulk DNA samples, and eDNA from water and sediment) for freshwater macroinvertebrate community assessment. This chapter particularly focused on the opportunities these new methods have to describe taxa for Water Framework Directive purposes.

Chapter 6: General Discussion

This chapter presents the overall findings for this thesis, including a short consultation on the use of eDNA by end-users, and details future applications of this work and potential limitations.

Chapter 2 Targeting the invaders – targeted detection of four priority freshwater invasive non-native species using environmental DNA

2.1 Abstract

Early detection is paramount for attempts to remove any invasive non-native species (INNS). Traditional methods rely on physical sampling and morphological identification, which can be both problematic when species are in low densities and/or are cryptic. Increasingly, the use of environmental DNA (eDNA) as a monitoring tool in freshwater systems is becoming an acceptable and widely used method for the detection of single species. Here we demonstrate the development and application of standard PCR primers for the detection of four INNS which are high priority for monitoring in the UK and elsewhere: Dreissenid mussels, *Dreissena rostriformis bugensis* (Andrusov, 1987) and *D. polymorpha* (Pallas, 1771), and freshwater shrimps, *Dikerogammarus villosus* (Sowinsky, 1984) and *D. haemobaphes* (Eichwald, 1843). We carried out a rigorous validation process for testing the new primers, including DNA detection and degradation rate experiments in mesocosm and a field comparison with traditional monitoring protocols. We successfully detected all four target INNS in mesocosm, but success was higher for mussels than shrimps. eDNA from single individuals of both mussel species could be detected within four hours of the start of the experiment. By contrast, shrimp were only consistently detected at higher densities (20 individuals), and this effect was greatest for *D. villosus*. In field trials, the two mussel species and *D. haemobaphes* were detected at all sites where the species are known to be present and eDNA consistently outperformed traditional kick sampling for species detection. However, *D. villosus* eDNA was only detected in one of five sites where the species was confirmed by kick sampling. These results demonstrate the applicability of standard PCR for eDNA detection of freshwater INNS, but also highlight differences between taxa in terms of the sensitivity of eDNA assays.

Note:

This chapter has been submitted to PeerJ for review

2.2 Introduction

Invasion biologists have highlighted the recent rapid increase in introductions, stating growth in human activity, namely tourism, trade and transport, as the driving factor over the last 25 years (Hulme, 2009; Gallardo and Aldridge 2013a). These studies and others have highlighted the influence of globalization (Sutherland et al., 2008), and risk of "invasional meltdown" (sensu Simberloff and Von Holl, 1999; Gallardo and Aldridge, 2014) as economic and population growth continues. Although geographically separated from continental Europe, the similar climatic conditions mean Invasive Non-Native Species (INNS) spreading through Europe, such as Ponto-Caspian species, are able to successfully establish in the UK if introduced. Concern over the increasing number of INNS has led to a number of horizon scanning studies aimed at identifying potential INNS through risk, threat and establishment opportunity (Gallardo and Aldridge 2013a; Gallardo and Aldridge; 2013b Roy et al., 2014). Roy et al., (2014) for example, concluded the potential impact, risk of arrival and risk of establishment of quagga mussels, *Dreissena rostriformis bugensis*, in the UK to be the highest out of 93 species examined. As predicted, the quagga mussel was first detected in the UK later the same year.

With growing pressure from legislators and limited funding to regulators to prevent further introductions of new INNS, efficient and effective monitoring tools are in high demand. Recent and rapid developments in molecular tools have meant a huge surge and investment in the use of DNA methods for biodiversity monitoring, in particular the use of environmental DNA (eDNA) (Lawson Handley, 2015). The first study to demonstrate this method successfully extracted and amplified DNA from the INNS American bullfrog, *Lithobates catesbeiana*. In this study, environmental DNA (eDNA) from pond samples was analysed using species specific primers and standard PCR. The method outperformed traditional monitoring approaches, producing reliable positive detections even when bullfrogs were present at low densities (Ficetola et al., 2008; Dejean et al., 2012). This bullfrog case study was revolutionary, and there soon followed a succession of similar studies utilising eDNA for the detection of a range of taxa in lentic, lotic and marine systems (Lawson Handley, 2015; Blackman et al., 2017b *in review*).

Unlike traditional monitoring methods, the successful detection of a species using eDNA does not rely on the collection of specimens (Ficetola et al., 2008). Instead, presence is determined by the detection of target DNA which is shed by the organism, for example in the form of sloughed cells, urine and faeces, or in decaying matter (Rees et al., 2014). The amount of DNA present in

the environment is influenced primarily by the species' DNA production rate, the degradation rate of the shed DNA, and the transport of DNA within the environment (Barnes et al., 2014; Barnes et al., 2015; Goldberg et al., 2015). Alongside development of species specific markers, studies have also focused on these areas with many concluding the availability of eDNA is highly dependent on the species being studied (Jerde et al., 2011; Thomsen et al., 2012b; Pilliod et al., 2013; Treguier et al., 2014; Roussel et al., 2015; Klymus et al., 2015; Jane et al., 2015) and the system in which they are present (Jane et al., 2015; Jerde et al., 2016; Shogren et al., 2017) and therefore any development of species specific primers should consider these variables.

Research within this field has moved at an incredibly fast pace, with over 60 INNS having species specific assays available to date (Blackman et al., 2017b *in review*); however, differing approaches have emerged. These can broadly be classified into "targeted" or "species specific" detection, and "passive" or "whole community level" approaches for detection. The approaches vary in their assay chemistry, technology, sensitivity and cost, and have advantages and disadvantages. The great majority of targeted assays have used either standard PCR (PCR) or probe-based real-time quantitative PCR (qPCR) - although droplet digital PCR (ddPCR) is an emerging technique that is growing in popularity. qPCR is often more sensitive than PCR for species detection (Thomsen et al., 2012a; Nathan et al., 2014) and high sensitivity makes this tool desirable for monitoring purposes. However, there is a trade-off between high sensitivity and a higher risk of false positives. For many applications, PCR is cheaper, less technically challenging, and more robust than qPCR, which may reduce the chance of false positives or negatives (Stoeckle et al., 2015; De Ventura et al., 2017).

In this study, we explored the potential application of standard PCR for the detection of key INNS in UK freshwaters. Four high priority species were targeted: quagga mussel (*Dreissena rostriformis bugensis*, Andrusov, 1897); zebra mussel (*Dreissena polymorpha*, Pallas 1871); killer shrimp (*Dikerogammarus villosus*, Sowinsky, 1894) and demon shrimp (*Dikerogammarus haemobaphes*, Eichwald 1841). These species all originate from the Ponto-Caspian area, and have spread rapidly throughout their invasive ranges via boat transportation, canals and river basin connections (Dick et al., 2002; Bij de Vaate et al., 2002; Timar and Phaneuf 2009). Both economic and ecological impacts are widely documented for all four species. The UK invasion history of these four species and the reasons for prioritising them for eDNA assay development, is discussed below.

Dreissenidae mussels:

D. polymorpha is widespread and common in the UK, having arrived in the 1820s potentially via the timber trade (Bij de Vaate et al., 2002; Quinn et al., 2014). *D. r. bugensis* is a much more recent invader, with the first UK record from 2014 in the River Wraysbury. Subsequent surveys showed the mussel was extensively distributed in the neighbouring reservoir; a facility used to supply drinking water and for leisure activities. This reservoir is subject to water transfers within the region and the mussel was subsequently found in neighbouring reservoirs. The monitoring of these two species within the UK poses a new challenge to regulators due their morphological similarity. The Dreissenid mussels both have huge impacts on ecosystem structure and function (Karatayev et al., 2007) and on the economy. For example, between US$161 - US$467 million was spent by water treatment and electric power facilities in North America on the control and removal of *D. polymorpha* between 1989- 2004 (Connelly et al., 2007).

The rapid spread and colonization of new waterbodies by Dreissenid mussels throughout the world has been aided by both human interaction and their unique ecology (Timar and Phaneuf 2009). Like other mussels, Dreissenids have a free floating planktonic veliger life stage, during which young can be dispersed over a large area downstream of parental populations (Ricciardi et al., 1995; Karatayev et al., 2002; Karatayev et al., 2015). Compared to many other mussel species, Dreissenids exhibit unique abilities to colonise new environments, by using protein-based byssal strands formed inside their shell to secure to hard surfaces, which can be a significant aid to transportation and establishment (Ricciardi et al., 1998; Karatayev et al., 2002; Aldridge et al., 2004; Timar and Phaneuf, 2009; Peyer et al., 2009). Colonization of new areas and establishment has been facilitated by the ability of Dreissenids to survive out of water for up to 15 days (Ricciardi et al., 1995) and tolerate a wide range of environmental extremes (Gallardo and Aldridge 2013b).

Monitoring and preventing the spread of *D. r. bugensis* is a priority within the UK, because of its recent arrival and potential to spread. The quagga mussel is likely to be able to invade a wider range of habitats than the zebra mussel, including areas with higher temperatures, lower rainfall, greater water depth, and lower dissolved oxygen (Nalepa et al., 2010; Quinn et al., 2014). *D. r. bugensis* are also able to spawn at lower temperatures than *D. polymorpha* (Roe and MacIsaac. 2011), which suggests they will potentially thrive in the cool UK climate. Both mussels are

described as “ecological engineers” (Karatayev et al., 2002; Karatayev et al., 2007; Roy et al., 2014), having influences on all trophic levels. In some instances, mussels provide increase in shelter and habitat for benthic macroinvertebrates (Karatayev et al., 2002), however they also compete for food and decrease diversity, and have been directly linked to declines in native Unionid mussels (Ricciardi et al., 1996). Dreissenid feeding behaviour also has negative effects on phytoplankton, and has been linked to greater numbers of cyanobacteria blooms (Karatayev et al., 2002).

Although previous studies have designed and tested primers for detection of quagga and zebra mussels, not all are suitable for discriminating between the two species (Peñarrubia et al., 2016). Others discriminate the species using a two-step PCR protocol, which was designed for tissue samples and may be less appropriate for eDNA due to its large amplicon size (Hoy et al., 2010). Studies by Mahon et al., (2011) and Egan et al., (2015), have both focused on detection of quagga mussel veligers in ballast water, using microfluidic chip and light transmission spectroscopy (LTM) technology, respectively. Recently, De Ventura et al., (2017) demonstrated the successful detection of eDNA from both species in the field with PCR and qPCR, using mitochondrial *COI* primers developed by Bronnenhuber and Wilson (2013). To our knowledge, no previous study has designed and tested Dreissenid species specific standard PCR primer pairs in controlled experiments to evaluate the rate of DNA production and probability of detection, nor evaluated their performance for detecting eDNA in the field against traditional methods for sampling.

Dikerogammarus species:

Dikerogammarus villosus and *haemobaphes* have spread in a similar way to the Dreissenid mussels; arriving in Germany by the late 1990s. *Dikerogammarus villosus* was first recorded in Grafham water, UK in September 2010 (MacNeil et al., 2010a), and to date its spread has been limited to only 5 further locations in the UK due to strict biosecurity measures (Check, Clean, Dry www.nonnativespecies.org). *Dikerogammarus haemobaphes* on the other hand, has rapidly colonized British waterways since its discovery in May 2012, spreading successfully through the river and canal networks. Both species are well documented as having significant negative effects on the macroinvertebrate community, particularly out competing native Gammaridae species (Dick et al., 2002; MacNeil et al., 2010a). *Dikerogammarus villosus* is particularly noted for its exceptional predatory capabilities (Dick et al., 2002; MacNeil et al., 2010a), its high reproductive

output (MacNeil et al., 2010a) and has been known to survive for up to 6 days out of water, allowing for extensive transportation on kit as well as in ballast water (Martens and Grabow 2008). Both *Dikerogammarus* species have been prioritised for monitoring by the UK Environment Agency, because of the potential for rapid spread, and high impacts on native fauna. To our knowledge, no species-specific primer pairs have been developed for *D. villosus* or *D. haemobaphes*.

The overall objective of this study was to develop and test targeted PCR eDNA assays for the four INNS named above. Our framework for developing and testing the assays, consisted of: 1. *in silico* and *in vitro* primer testing; 2. single species mesocosm experiments to evaluate eDNA detection probability over time at three different densities (1, 5 and 20 individuals) and rates of eDNA degradation; and 3. testing the efficiency of the targeted PCR eDNA assays compared to traditional kick-net sampling in the field.

2.3 Methods

2.3.1 Specimen sampling and tissue DNA extraction

Specimens of all four target INNS taxa were collected at sites with known populations, two weeks prior to the beginning of each mesocosm experiment. These sites were as follows: *D. r. bugensis*, Wraysbury River, UK Grid. Ref. TQ 02680 73204; *D. polymorpha* and *D. haemobaphes*, Rutland Water, SK92956 05963; *D. villosus*, Grafham Water, TL 15081 67289. Specimens were kept in tanks with continuous aeration and fed dried *Cyclotella* and leaf material *ad libitum*. Samples from the most closely related native taxa were also collected for tissue DNA extraction and primer testing (*Gammarus fossarum/pulex, Crangonyx pseudogracalis, Sphaerium corneum* and *Anadonta anatina*). Tissue samples from four individuals of each INNS or native species were extracted using the DNeasy Blood and Tissue Kit® (Qiagen, Hilden, Germany) according to the manufacturer's protocol.

2.3.2 Species specific primer development

For all four INNS, species specific primers were designed and tested in silico with Primer BLAST (Ye et al., 2012) using all available *COI* reference sequences from GenBank (*D. r. bugensis* – 7 sequences, *D. polymorpha* – 31 sequences, *D. villosus* – 22 sequences and *D. haemobaphes* – 7

sequences, see Supplementary Information I: Table S1 for details of GenBank accession numbers). Thirty-eight primer pairs were designed and tested *in silico*. Twenty-three primer pairs were then selected for *in vitro* testing based on low self-complementarity, no unintended target species amplification and small target length. Primers were preferentially selected with an amplicon size < 200 bp, to be suitable for amplification of degraded eDNA (Deagle et al., 2006; Jerde et al., 2011; Bronnenhuber and Wilson 2013; Mächler et al., 2014; Ardura et al., 2015a). Details of all primer pair tests can be found in Supplementary Information I Table S2.2. The 23 primer pairs were tested *in vitro* on tissue samples of target INNS and three related non-target taxa i.e. the congeneric INNS and two native taxa which are likely to co-occur in the same habitat. Serial dilutions of neat tissue-extracted DNA (3-5 ng/µl) to 1:1000 dilutions (0.003-0.005 ng/µl) were carried out to establish the Limits of Detection (LoD) for each primer pair (Supplementary Information I, Table S2.3). PCRs were carried out in 25 µl volumes with MyTaq Red Mix (Bioline, UK) containing: 10 µM of each primer and 2 µl of undiluted DNA template. PCRs were performed on an Applied Biosystems Veriti Thermal Cycler with the following profile: initial denaturation at 94°C for 3 min, followed by 37 cycles of denaturation at 94°C for 30s, annealing at 65°C for 1 min and extension at 72°C for 1 min 30s, with a final extension time of 10 min at 72°C. PCR products from tissue samples were visualised by gel electrophoresis and stained with GelRed (Cambridge Bioscience Ltd, UK). Four PCR products per species were Sanger sequenced by Macrogen Europe in the forward direction. All sequences were compared with the National Centre for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) using BLAST to confirm species identification.

2.3.3 Mesocosm experiments

Mesocosm experiments were carried out from January to August 2016 to test the sensitivity of the selected primer pairs under controlled densities. Each experiment was conducted in 15 L plastic tanks with fitted lids. Tanks were located in a climate controlled facility, where temperature averaged 16°C (range 14-18°C) with light:dark cycles of 16 h:8 h . All tanks, aeration equipment and sampling equipment was sterilized in 10% commercial bleach solution for 10 minutes, then rinsed with 10% MicroSol detergent (Anachem, UK) and purified water prior to the experiment. Sampling and filtering equipment was also cleaned using the above method between each sampling event. Each tank was filled with water collected from Hotham Beck (SE 89133 32489) which has no previous records of the four target INNS. Tanks were supplied with constant air via sterile tubing and aeration stones for 48 hours prior to the start of the experiment and covered for the duration of the experiment with a fitted lid.

For each species, the experiment consisted of 10 tanks, representing three replicates of three treatment densities (1, 5 and 20 individuals) and a control tank with no individuals. Specimens of similar total biomass were used in the density replicates, in order to minimise any influence of different biomass (see Supplementary Information I, Tables S2.4, S2.6, S2.8 and S2.10 for biomass information gathered pre- and post-mesocosm experiments). Room temperature, control tank and water temperature was recorded prior to each sampling event (see Supplementary Information I, Tables S2.5, S2.7, S2.9 and S2.11 for temperature measurements taken during the mesocosm experiments). Tank water temperature was kept below 10°C for both Dreissenidae mesocosm experiments to minimise any potential spawning events. Before the experiments, all specimens were examined to confirm species identification, any *Dikerogammarus* females carrying eggs or juveniles were excluded from the experiment to avoid influencing the DNA concentration. Before the specimens were added to the tanks, a water sample was collected and filtered to ensure no contamination from the target taxa; this sample was recorded as 0 hours. Tanks were sampled over 42 days at 4hrs, 8hrs, 24hrs, 7 days, 15 days and 21 days with the species present to investigate the probability of eDNA detection over time and at different densities. On day 21, the specimens were removed from the tanks and sampling continued at 22 days, 28 days, and 42 days to document the rate of DNA degradation. A total sample size of N = 100 was collected per species.

For each sampling event, the tank water was homogenised by stirring with a sterile spatula before collecting 200 ml water from each tank. Samples were vacuum filtered through sterile 47 mm diameter 0.45 μm cellulose nitrate membrane filters with pads (Whatman, GE Healthcare, UK) immediately after collection, using Nalgene filtration units (Thermo Fisher Scientific) in combination with a vacuum pump (15~20 in. Hg, Pall Corporation) in a dedicated eDNA laboratory at the University of Hull, UK. Filter papers were then placed in sterile petri dishes, sealed with parafilm and stored at -20 °C until extraction. The filtered water was then returned to the tank to maintain the water volume. This process was completed within one hour. The filtration units were cleaned with 10% commercial bleach solution and 5% MicroSol, and then rinsed thoroughly with deionized water after each filtration to prevent cross-contamination. All DNA extractions were carried out using a protocol modified from Bolaski et al., (2008) (for the full extraction protocol, see Supplementary Information II). Mesocosm samples were PCR amplified using the species-specific primers and conditions previously described. PCR products were then visualised on a 1.5% agarose gel stained with GelRed (Cambridge Bioscience Ltd, UK). Three

PCR products from each species/mesocosm experiment were sequenced to confirm primer specificity (Macrogen Europe).

2.3.4 Field trials

Water samples were collected at sites with previous records of the target INNS to test and verify the efficiency of each INNS assay in the field. For *D. r. bugensis, D. polymorpha* and *D. haemobaphes*, three UK Environment Agency macroinvertebrate monitoring sites were selected on three river catchments (Colne, Welland and Nene) (n = 9) (Fig. 2.1 A-C). Six samples were collected at each of the three sites, each sample consisted of 3 x 500 ml (n = 54 per INNS). Field samples for *D. villosus* were collected from the shoreline of the Grafham Water reservoir using the same protocol as for the lotic samples (3 sites x 3 x 500 ml replicates, hence n = 9, Fig. 2.1D). Each 500 ml sub-sample was filtered and extracted independently. For *D. villosus*, additional single 2 L water samples from Wroxham Broad and Pitsford Water (Fig. 2.1E and F respectively), collected during a different study, were also tested. Sites were surveyed after eDNA sample collection using standard 3-minute kick samples (Murray-Bligh, 1999). Sample bottles filled with ddH_2O were taken into the field as sample blanks. Samples were processed within 24 hours using the same method as the mesocosm samples. Each technical replicate was PCR amplified three times to avoid false negatives. Four PCR products from each mesocsom/field sample were sequenced to confirm primer specificity.

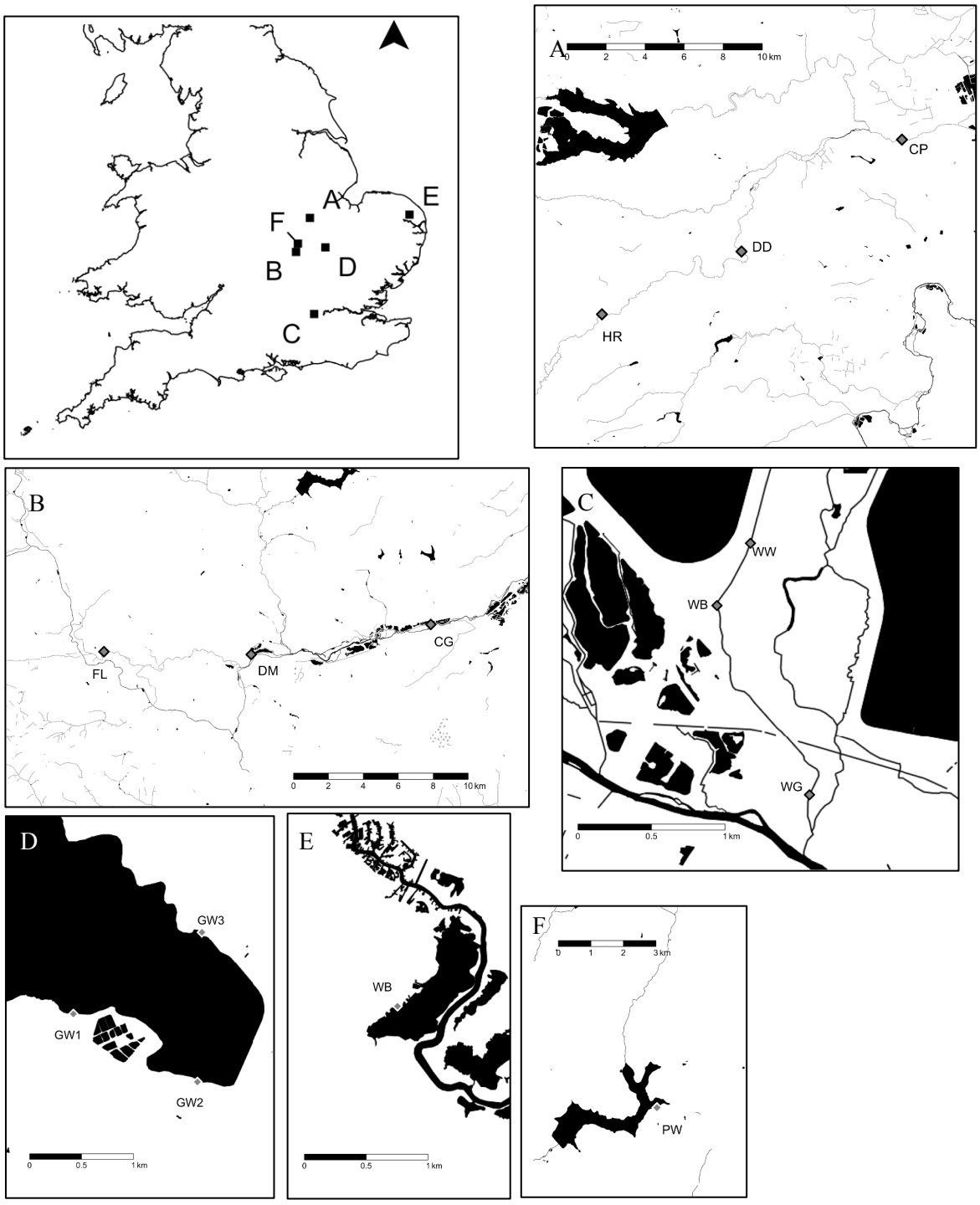


Figure 2.1 Site locations for field trials. All sample sites are referred to from upstream to downstream, waterbodies are in black, sample points are marked with a diamond. A - Sampling sites from *D. r. bugensis* were on the River Wraysbury at Wraysbury weir (WW), Wraysbury bridge (WB) and Wraysbury Gardens (WG). B - Sample sites for *D. polymorpha* were on the R. Welland at Harrington (HR), Duddington (DD) and Copthill (CP). C - Samples sites for *D. haemobaphes* were on the R. Nene at Flore's Road Bridge (FR), Duston Mill (DM) and Cogenhoe (CG). D, E, F - Sample sites for *D. villosus* were carried out on three reservoirs: Grafham Water (GW1, GW2, GW3), Wroxhom Broad (WB) and Pitsford Water (PW).

2.3.5 Data Analysis

Binomial Generalized Linear Models (GLMs) with a logit link function were used to investigate the influence of density or total biomass and time since the start of the experiment (until the taxa were removed from the mesocosm) on the probability of detection. Models were checked by testing whether the residual deviance fitted a chi squared distribution. All data analyses were performed in R v.3.3.1. (R Core Team 2017), with GLMs performed using the MASS package (Venables et al., 2002).

2.4 Results

2.4.1 Primer specificity

Of the twenty-three primer pairs selected after *in silico* and *in vitro* testing, four (one for each species) were selected based on our criteria of good target amplification with no cross-amplification of non-target species (DRB1, DP1, DV1 and DH2, see Table 2.1). The *D. r. bugensis* primer pair, DRB1, amplified 29 published *D. rostriformis, D. bugensis* and *D. rostriformis bugensis* sequences in silico with no mismatches. The *D. polymorpha* primer pair, DP1, amplified 45 published *D. polymorpha* and subspecies (*D. p. polymorpha, D. p. gallandi and D. p. anatolia*) in silico. Of the published *D. polymorpha* sequences, one had a mismatch in the forward primer (Accession number AF510508) and a second sequence had two mismatches in the forward and one in the reverse primer (Accession number JQ435817). Note that the forward primer pair selected for *D. polymorpha* shares a 16 bp overlap with DpoCOI3F designed by Bronnenhuber and Wilson (2013) but our primer pair, DP1, amplifies a much shorter sequence (73 bp). *The D. villosus* primer pair, DV1, amplified 23 *D. villosus* sequences in silico. Two published sequences of *D. villosus* from the Ukraine had mismatches to our primer pair; 1 mismatch with the forward primer and two in the reverse (Accession numbers KM208873 and EF570297). Finally, the *D. haemobaphes* primer pair, DH1, amplified 7 published sequences in silico. Three of these sequences are non-target species of marine gastropod: *Thuridilla albopustulosa* (Accession number KM086443), *Hemicycla pouchadan* (Accession number GU598226), *Caucasotachea calligera* (Accession number KT794407). Since these are marine species and currently not recorded in the UK, it is unlikely they will generate false positives in our tests but this should be considered for wider applications. (Table See Supplementary Information I Table S2.12 for full species specific primer mismatches)

Table 2.1. Primer pairs designed for this study and used for the detection of 4 target INNS.

Target species	Primer	Primer sequence	Amplicon length (bp)
Dreissena rostriformis bugensis	DRB1_F	GGAAACTGGTTGGTCCCGAT	188
	DRB1_R	GGCCCTGAATGCCCCATAAT	
Dreissena polymorpha	DP1_F	TAGAGCTAAGGGCACCTGGAA	73
	DP1_R	AGCCCATGAGTGGTGACAAT	
Dikerogammarus villosus	DV1_F	TCTTGGCAGGTGCCATTACG	87
	DV1-R	GAATAGGATCACCCCCGCCT	
Dikerogammarus haemobaphes	DH2_F	TAGGTCACAGGGGTGCTTCT	295
	DH2_R	AAGTGCTGGTAAAGAATAGGA TCT	

Species specific primer testing on target tissue samples yielded positive PCR amplification of a single band at the expected size for all four species (Fig. 2.2). The LoD for DRB1, DP1, and DH1 primer pairs was ~0.005 ng/µl DNA per reaction (1:1000 dilutions of neat tissue DNA, Fig. 2.2 A, B, D). For DV1, the LoD was a 1:100 dilution, which corresponds to approximately 0.03 ng/µl of target DNA (Fig. 2.2 C). No bands of the expected size were obtained in the cross amplification tests; however, much larger, non-specific bands, were seen in non-target species amplifications for DP1, and DV1 (Fig. 2.2 B and C). Due to the substantial size difference these non-specific bands are easily distinguishable from and the target band size and thus not lead to false positive detections. Sequences generated from PCR products from all tissue, mesocosm and field samples were verified as being from the correct target species. Some sequences generated from *D. villosus* and *D. polymorpha* were of poor quality due to the short amplicon length, but sequences of 20 – 37 bp were matched with 100% success.

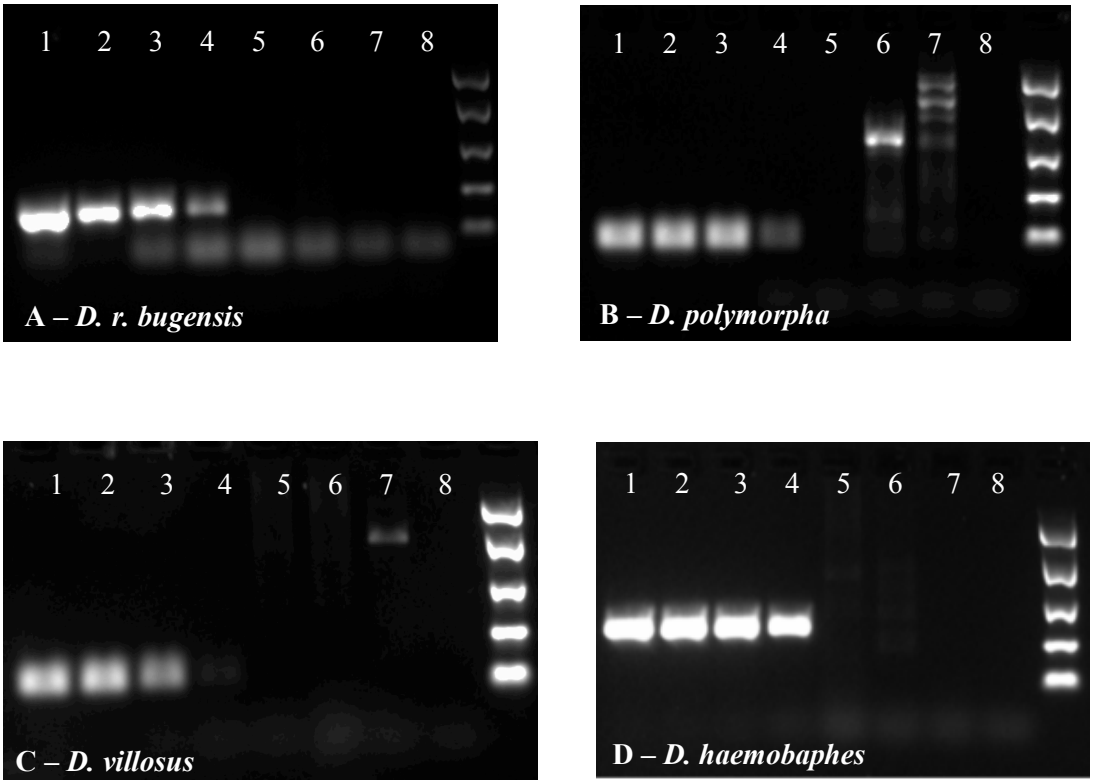


Figure 2.2: Results of in vitro primer testing. A – *Dreissena rostriformis bugensis* (primer pair DRB1), B – *Dreissena polymorpha* (DP1), C – *Dikerogammarus villosus* (DV1) and D – *Dikerogammarus haemobaphes* (DH2). Lane 1 contains undiluted target INNS tissue DNA (3-5 ng/µl per reaction), lanes 2-4 contain a dilution series of the target tissue (lane 2 1:10 dilution, ~0.3-0.5 ng/µl per reaction; lane 3 1:100 dilution, ~0.03-0.05 ng/µl per reaction; lane 4 1:1000 dilution, ~0.003-0.005 ng/µl per reaction). Lanes 5 and 6 contain closely related native species found in the UK: for the Dreissenid mussels (A and B): *Anadonta anatina* and *Sphaerium corneum*, and for the *Dikerogammarus* species (C and D): *Gammarus fossarum/pulex*, and *Crangonyx pseudogracalis*. Lane 7 contains the paired INNS and lane 8 is a PCR negative (ddH_2O). The final lane is DNA EasyLadder I (Bioline, UK) with fragment sizes of 100 bp, 250 bp, 500 bp, 1000 bp and 2000 bp.

2.4.2 Mesocosm experiments

All tanks were clear of any target DNA prior to adding the target species (i.e. at time = 0). No contamination was detected in any of the control tanks over the course of the experiments (see Supplementary Information III, Fig. S2.1, S2.2, S2.3 and S2.4 for gel images of all mesocosm samples).

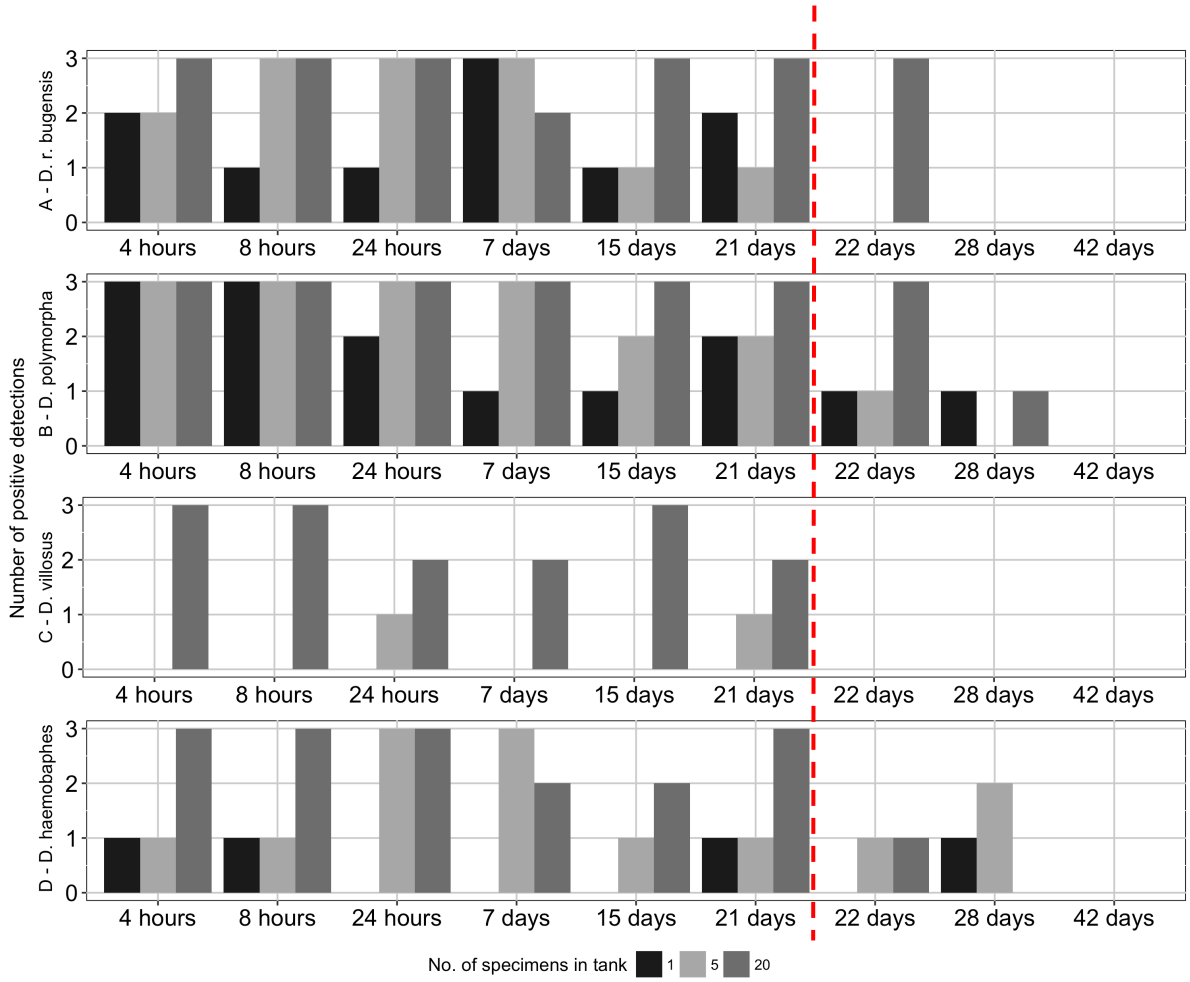


Figure 2.3: Species detection in mesocosm experiments: each graph indicates the number of positive detections from three replicates taken from each treatment (specimen density) during the 42 day experiment. Specimens were removed after 21 days (indicated by the dashed red line). A – *Dreissena rostriformis bugensis*, B - *Dreissena polymorpha*, C - *Dikerogammarus villosus* and D - *Dikerogammarus haemobaphes*.

Both Dreissenid mussel primers showed positive detection of their target species in all three replicates at the four hour sampling event (Fig. 2.3A and B). At least one positive replicate was

obtained for every sampling point over the first 21 days. For *D. r. bugensis*, density and total biomass significantly influenced the probability of detection by standard PCR. Of these two measures, total biomass was the more significant predictor in GLMs and generated the lowest Akaike Information Criterion (AIC) (GLM, z = 2.054, P = 0.040, AIC 54.926). Time since the start of the experiment was not a significant predictor of detection probability. After removal of *D. r. bugensis*, DNA was only detected in tanks with the highest mussel density (20) 24 hours after removal. DNA from these tanks was no longer detected at day 28 (7 days after removal). For *D. polymorpha*, both time and density (but not total biomass) were significant predictors of the detection probability. Of the two measures, density was the strongest predictor with the lowest AIC (GLM, z = 1.969, P = 0.0490, AIC 32.823). DNA from *D. polymorpha* persisted to day 28 in two of the three density treatments. No positive detections were made for any density treatment on day 42 (21 days after removal) (see Supplementary Information I for GLM outputs).

Dikerogammarus species: DNA from both target species was consistently detected at each sampling point, between 4 hours and 21 days for the 20 individual density treatment (Fig. 2.3 C and D). *Dikerogammarus haemobaphes* was also consistently detected in the 5 individual density treatment, and after 4, 8 hours and 21 days in the single individual treatment (Fig. 2.3D) *Dikerogammarus villosus* was only detected at two sampling points (after one and 21 days) in the 5 individual density treatments and was not detected at all in the single density treatment (Fig. 2.3C). *D. villosus* and *D. haemobaphes* primers amplified non-specific DNA during the experiment. However, these non-specific band were of substantially different to the target band size in both cases (*D. villosus* non-target bands > 300 bp, *D. haemobaphes* non-target band size < 100 bp) (See Supplementary Information III Fig. S2.3 and S2.4 for agarose gels from each *Dikerogammarus* mesocosm experiment). Density or total biomass significantly influenced the probability of detection for both shrimp species. Total biomass was the most significant predictor for D. *villosus* and had the lowest AIC (GLM, z = 4.491, P = 0.00000709, AIC: 29.285). Density was the most significant predictor for *D. haemobaphes* (GLM, z = 3.407, P = 0.000658, AIC: 59.887). Time since the start of the experiment was not a significant predictor of detection probability for either species. Positive detection of *D. villosus* at all densities after removal at any density was lost within 24 hours of the specimens being removed, however detection of *D. haemobaphes* continued until 7 days after removal.

2.4.3 Field trials

Dreissenid mussels: *D. r. bugensis* specimens were found by kick-sampling at all sites surveyed but the number of individuals found decreased with distance along the River Wraysbury, from the main source population at Wraysbury Reservoir (Fig. 2.4A). Detection probability by kick-sampling was 33% (6 samples out of 18). Positive eDNA detections were obtained for every sampling replicate at each of the three sites along the River Wraysbury, hence eDNA detection probability is 100% (Fig. 2.4A and Supplementary Information III Fig. S2.5). *D. polymorpha* was found by kick-sampling in only one of three sites (Duddington, Fig. 2.4B) although the species is known to be present throughout the sampled catchment. Detection probability for kick-sampling was 11% (2/18 samples). Positive eDNA detections for *D. polymorpha* were obtained in 88.9% (16/18) of samples and 54% (29/54) of the PCR replicates, including in sites where specimens of *D. polymorpha* were not found (Fig. 2.4B and Supplementary Information III Fig. S2.6).

Dikerogammarus species: *D. haemobaphes* was detected in all kick-samples at two of the three sites sampled (detection probability 66.7%, Fig. 2.4 D). Positive eDNA detections for *D. haemobaphes* were obtained in 83.3% (15/18) of samples and 59.3% (32/54) of PCR replicates (Fig. 2.4D). Positive detections were obtained for three of the six samples at Flore Road Bridge, where the species was not detected by kick-sampling (Fig. 2.3 D and Supplementary Information III, Fig. S2.8). *D. villosus* was detected in all five kick samples obtained (detection probability 100%) at varying density (Fig. 2.4C). However, the species was only detected in one of the 5 samples using eDNA (detection probability 20%, Fig. 2.4C and Supplementary Information III S7). The *Dikerogammarus* primer pairs were more challenging to employ in the field testing than the *Dreissena* primers, with weaker target bands and more non-specific amplification (Supplementary Information III, Figs. S2.7-8).

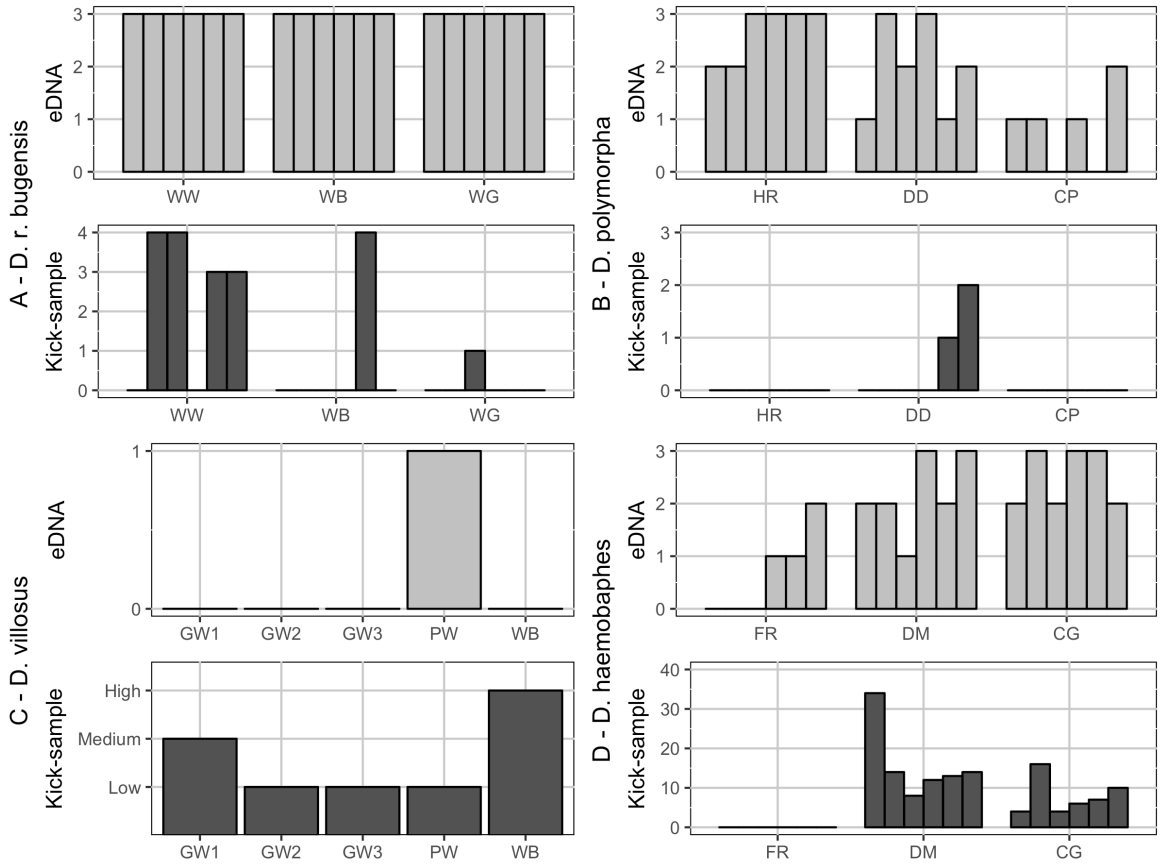


Figure 2.4: Species detection in field experiments. Each graph indicates the number of positive detections (PCR replicates for eDNA and specimens collected from kick-samples) from each field sample at each site. A – *Dreissena rostriformis bugensis* – River Wraysbury, B - *Dreissena polymorpha* – River Welland and C – *D. villosus* - Grafham Water, Pitsford Water and Wroxham Broad and D - *Dikerogammarus haemobaphes* – River Nene.

2.5 Discussion

Rapid, cost-effective tools are needed for detection of newly invading, or spreading invasive non-native species. Here, we designed and tested PCR primer pairs for four INNS: *D. r. bugensis, D. polymorpha, D. villosus* and *D. haemobaphes*, which are high priority for monitoring in the UK and beyond. Primers were tested in silico and in vitro, then in a series of mesocosm experiments and field trials. The four INNS primer pairs amplify target tissue at a low concentration (0.005-0.03 ng/ul) which is in line with other eDNA species specific primer assays (e.g. Ardura et al., 2015a), with no cross-species amplification with closely related or native species present in the UK. All four species were detected from eDNA collected from water samples in both laboratory and field trials.

eDNA could be detected in mesocosms within 4 hours of the start of the experiment, and with the exception of *D. villosus*, detection at this first time point was possible from just one individual. Dreissenid eDNA was detected at every sampling point at all three densities in the mesocosms, and greatly outperformed kick-sampling for detection in the field. Detection of *Dikerogammarus* eDNA was more challenging in both the mesocosm and field experiments but both species were consistently detected in the mesocosms at high density (20 individuals) and *D. haemobaphes* was also consistently detected at medium density (5 individuals). Field detection was higher for eDNA than kick-sampling for *D. haemobaphes* but detection of *D. villosus* was more variable. Below we highlight the range of factors that likely interact to determine the success of eDNA detection in real-world applications.

2.5.1 Mesocosm trials

Mesocosm experiments have been advocated (De Ventura et al., 2016), and performed by previous studies (Dejean et al., 2011; Thomsen et al., 2012b; Sansom and Sassoubre., 2017) to allow information to be gathered, such as species-specific DNA production rates, persistence and degradation over time. These, in turn, will inform regulators if this method is appropriate for the detection of target taxa. Here both abundance variables (density and total biomass) were significant predictors of detection for all four species. Hence, there is a positive relationship between abundance and detection, as found in previous studies (e.g. Thomsen et al., 2012b). Mesocosm experiments also demonstrated the rapid depletion of DNA once the specimens had been removed, with no detectability 7 days after removal, also in agreement with similar studies (Dejean et al., 2011; Thomsen et al., 2012b). However, there were differences between species in terms of detectability and DNA did not accumulate in a linear fashion over time, as discussed below.

The mesocosm experiments performed in this study were useful for determining the assay sensitivity and for identifying differences in detectability between species. Our mesocosm experiments revealed that the *Dreissena* primers are highly sensitive and robust – being able to detect single individuals within four hours and then consistently throughout the course of the experiment. The *Dikerogammarus* assays were less sensitive than those for the mussels, but *D. haemobaphes* was still consistently detected at medium and high densities, and *D. villosus* consistently detected in the highest density treatment. This likely reflects both differences

between the assays in terms of primer robustness and physiological differences between the two species pairs. The high success for Dreissenid mussels is likely due, at least in part, to the fact they were able to continuously filter feed on algae and phytoplankton present in the water column during our experiments, as they would in the wild, enabling them to maintain an active metabolism and eDNA production via sloughed cells, faeces and pseudofaeces (Sansom and Sassoubre., 2017). By contrast, *Dikerogammarus* metabolism may have been limited by the availability of only phytoplankton and algae as their diet in the wild is much more varied. Furthermore, no evidence of moulting was found during the course of the *Dikerogammarus* experiment, which is likely to be a main source of eDNA in the wild. Previous studies have suggested that organisms with exoskeletons (such as Crustacea) can be hard to detect with eDNA, potentially due to low shedding rates (Treguier et al., 2014, Dunn et al., 2017). It is clear from our study and others that DNA production by a species and its availability in the water can vary substantially between different species.

Although differences in species physiology may explain the differences in detection of the Dreissenid and *Dikerogammarus* species pairs, it does not explain differences within pairs, where anatomy and physiology are very similar. This difference in eDNA detection was most acute for *D. villosus* and *D. haemobaphes*. The detection of *D. haemobaphes* but not *D. villosus* in the single density treatment could at least partly be explained by difference in biomass (means 0.97 g and 0.13 g respectively). This explanation is less likely to account for differences in detection in the five individuals density treatment since biomass was more similar for the two species (means *D. haemobaphes* 0.69 g and *D. villosus* 0.50 g). Higher sensitivity and/or robustness of the *D. haemobaphes* primer pair, is likely an important contributing factor.

We might expect that as long as DNA production rate is greater than the degradation rate, (as seen in models produced by Thomsen et al., (2012b), eDNA availability should increase over the course of the experiment. Under this prediction, we expect the number of positive detections to increase over time, and for there to be an interaction with density. However, in our experiments eDNA availability fluctuated rather than accumulated over time. Of the four mesocosm experiments, only *D. polymorpha* showed a significant correlation with time. The fluctuation in detection rate may in part be due to activity of the specimens. In the case of the Dreissenid mussels, although not seen as strongly with the *D. polymorpha*, filter feeding is both a source of DNA but may also be a mechanism to ingest it. Similarly, activity by the shrimp may have reduced

over the course of the experiment with reduction of food resources in the water column; however, these effects are unlikely to hinder detection in the field.

Inhibition may also be a cause of fluctuation in detection rate, and has been reported in other mesocosm tests (Sassoubre et al., 2016) and field trials (Jane et al., 2015). We do see an increase in band strength in both Dreissenid mesocosm experiments between 4 and 24 hours, and for the high density *Dikerogammarus* mesocosm tanks followed by failed amplification and lower band strengths, however this is unlikely to be due to inhibition as we do not see any consistency in failed amplification in individual tanks. Jane et al., (2015) demonstrated a link between PCR inhibition and seasonality which should be taken into account when applying these methods to field sampling.

2.5.2 Field application

For three INNS, *D. r. bugensis, D. polymorpha* and *D. haemobaphes*, the field tests successfully detected the target species in a high proportion of samples. For these species, eDNA consistently outperformed traditional kick-sampling, including *D. r. bugensis* detection in every replicate of every sample, compared to 33% of kick samples. Below we discuss the reasons for the discrepancies between eDNA and kick-samples for all four species.

There are numerous influences on the persistence of eDNA in rivers that have been well documented (Deiner and Altermatt, 2014; Barnes et al., 2014; Jane et al., 2015; Jerde et al., 2016; Shogren et al., 2017). In our study, the higher detection from eDNA samples than kick-samples at sites downstream implies that DNA is not restricted or localized, rather it is being transported from upstream sources, which is in agreement with other studies (Deiner and Altermatt, 2014; Jane et al., 2015). However, to what extent the DNA is being transported is still largely unknown. Previous work on river morphology states substrate type and the related flow regime, are huge influences on DNA transportation, retention and subsequent resuspension (Shogren et al., 2017; Jerde et al., 2016). Finer substrates (i.e. pebbles, gravel) are likely to retain DNA for longer and absorb it quicker than larger substrates (Cobbles) (Shogren et al., 2017; Jerde et al., 2016). However, Shogren et al., (2017) also showed stochastic variation in the retention of DNA by these differing substrates sizes with no clear consensus. In our study catchments, we have a diverse mixture of substrate sizes, including sediment which will readily retain DNA for long periods of

time (Turner et al., 2015). Therefore, we are unable to accurately gauge how far eDNA may be transported.

The samples collected for *D. r. bugensis* followed a density gradient along the length of the river (2 km), but due to higher flows during our survey following rainfall, we do not see a decrease in band strength on the agarose gels, or decrease in detection rate, which is consistent with previous research (Jane et al., 2015). This is likely due to transportation of DNA as previously mentioned, and also increased water mixing. We know that eDNA is not uniformly distributed through a river (Macher and Leese, 2017), therefore an increase in flow is likely to increase the dispersion of eDNA in a waterbody through mixing (Shogren et al., 2017). We therefore see a greater number of positive detections as observed with *D. r. bugensis*. Similarly, we see variation in the eDNA detection for both D. *polymopha* and *D. haemobaphes*, due to the relatively lower flows during these surveys which have caused a reduction eDNA distribution across the river. However, we must also note that although we have a fluctuation in the number of detections across the samples at each site for these species, we do still see evidence of eDNA being transported down the catchment. For both *D. polymorpha* and *D. haemobaphes*, we detect DNA at sites where they were not physically collected. This is not a false positive result but rather a detection of specimens from discrete populations further up the catchment. This greater variability in detection due to the lower flow conditions is likely to demonstrate the true variation encountered when surveying lotic systems for target INNS.

eDNA samples for *D. villosus* were obtained from three lentic water bodies where the species is known to be present in differing densities. In line with previous work that has demonstrated spatial structuring of eDNA in lentic water bodies (Hänfling et al., 2016), we predicted eDNA detection to be more representative of the community in the immediate proximity. In contrast to our expectations, *D. villosus* was only detected by eDNA at one of five sites, despite detection at all five sites with traditional sampling. Reasons for the greater difficulty detecting *D. villosus* in the wild maybe due to the lower DNA shedding rates and poorer assay performance as discussed above. Combined with localisation of eDNA, this suggests that a greater sampling resolution may be required to detect *D. villosus* and other species with low shedding rates in lentic water bodies.

2.6 Conclusion

The development of species specific primers like those demonstrated in this study show a much needed advancement in the early detection of INNS. This approach has several applications for monitoring high risk pathways (such as ballast water and water sport equipment) and enabling the accurate mapping of current INNS spread within the UK if a new invasion occurs, which in turn will aid biosecurity measures. Importantly, using PCR primers allows a simple method for those with monitoring responsibilities to utilise eDNA sampling without significant investment in qPCR, ddPCR or Next Generation Sequencing. With current methodology, eradication is often not an option, particularly in flowing water systems. However, early detection of species, such as *D. villosus* and *D. r. bugensis* which are known to have first been introduced to the UK in contained systems (e.g. reservoirs), by using species specific primers and effective eDNA surveillance methods may have allowed earlier implementation of containment procedures and significantly slowed their spread. It is our hope that this study demonstrates an efficient and cost-effective framework for the development of PCR primers for target INNS, which can be used by regulatory bodies with responsibility for INNS monitoring.

2.7 Supplementary Information I

Table S2.1. GenBank accession numbers used for designing species specific primers for INNS target taxa.

Species	GenBank Accession number
Dreissena polymorpha	JX099437, EF414495, HM210079, EU484444, AF492005, KY091877, KX537632. KC429149, AM746677, EU484441, EU484431, EU414494, AF120663, AF474404, EU484433. EF414493, AM748999, EU484435, AM748975, EU484448, AF479636, U47653, AM749001, AM7489900, AM748988, AM748996, AM748992, AM748989, AM748997, JQ771951, DQ840122
Dreissena rostriformis bugensis	EF080862, JX945980, EF080861, U47651, AF495877, EU484436, KJ881409
Dikerogammarus villosus	KP814111, KM208862, KM208874, KM208872, KM208866, KM208865, KF478540, KF478533, KM208871, KM208870, KM208869, KM208868, KM208867, KM208863, KM208873, KM208864, AY529048, KT075266, KT075265, KT075264, KF478581, EF570297
Dikerogammaurs haemobaphes	KT075268, KT075267, KM009057, AJ44092, AJ440920, AJ440919, AY529049

Table S2.2: Primers designed and tested for species specific amplification of four target INNS

Tested	Primer name	Primer direction	Sequence (5'->3')	GC%	Self complementarity	3' Self complementarity	Amplicon	Species amplified
in silico	DVIL5	Forward primer	CATAGCGGTGCTTCCGTTGA	55.00	3.00	1.00	275	*Dikerogammarus villosus* *Turcogammarus spandli* *Allognathus hispanicus*
		Reverse primer	ATAGGATCACCCCCGCCTC	63.16	4.00	0.00		
in silico	DVIL6	Forward primer	CTTGGCAGGTGCCATTACGA	55.00	8.00	0.00	85	*Dikerogammarus villosus* *Hyalella azteca*
		Reverse primer	AATAGGATCACCCCCGCCTC	60.00	4.00	0.00		
in silico/in vitro	DVIL3	Forward primer	CGGTGCTTCCGTTGATCTTG	55.00	4.00	0.00	270	*Dikerogammarus villosus* *Petaloconchus keenae*
		Reverse primer	ATAGGATCACCCCCGCCTC	63.16	4.00	0.00		
in silico/in vitro	DVIL4	Forward primer	ATAGGCCATAGCGGTGCTTC	55.00	4.00	0.00	282	*Dikerogammarus villosus* *Turcogammarus spandli* *Setobaudinia herculea*
		Reverse primer	AATAGGATCACCCCCGCCT	57.89	4.00	0.00		
Selected!	DVIL1	Forward primer	TCTTGGCAGGTGCCATTACG	55.00	8.00	2.00	87	*Dikerogammarus villosus*

		Reverse primer	GAATAGGATCACCCCCGCCT	60.00	4.00	0.00		
in silico/in vitro	DVIL2	Forward primer	TGGCAGGTGCCATTACGATA	50.00	8.00	2.00	81	*Dikerogammarus villosus*
		Reverse primer	TAGGATCACCCCCGCCTC	66.67	4.00	0.00		
in silico/in vitro	DHAE1	Forward primer	CATAGGTCACAGGGGTGCTT	55.00	3.00	0.00	297	*Dikerogammarus haemobaphes*
		Reverse primer	AAGTGCTGGTAAAGAATAGGATCTC	40.00	4.00	2.00		
Selected!	DHAE2	Forward primer	TAGGTCACAGGGGTGCTTCT	55.00	3.00	0.00	295	*Dikerogammarus haemobaphes* *Gyliotrachela hungerfordiana* *Setobaudinia gumalamala* *Hemicycla pouchadan* *Thuridilla albopustulosa*
		Reverse primer	AAGTGCTGGTAAAGAATAGGATCT	37.50	4.00	2.00		
in silico/in vitro	DHAE3	Forward primer	GTCACAGGGGTGCTTCTGTT	55.00	4.00	2.00	292	*Dikerogammarus haemobaphes* *Hemicycla pouchadan* *Ligia Costasiella ocellifera* *Anguispira alternata* *Allognathus hispanicus campanyonii* *Actia nr. cinerea* *Hemicycla inutilis* *Hemicycla modesta* *Hemicycla incisogranulata* *Gyliotrachela hungerfordiana*
		Reverse primer	AAGTGCTGGTAAAGAATAGGATCTC	40.00	4.00	2.00		

in silico/in vitro	DHAE4	Forward primer	CACAGGGGTGCTTCTGTTGA	55.00	4.00	1.00	291	*Dikerogammarus haemobaphes* *Mongoloniscus sinensis* *Ligia occidentalis* *Phoridae sp* *Fleutiauxellus curatus* *Tanytarsus unagiseptimus* *Haloniscus sp.* *Ligia sp* *Phoxocephalidae sp*
		Reverse primer	TAAGTGCTGGTAAAGAATAGGATCT	36.00	4.00	2.00		
Selected!	DPOL1	Forward primer	TAGAGCTAAGGGCACCTGGAA	52.38	4.00	0.00	73	*D.polymorpha*
		Reverse primer	AGCCCATGAGTGGTGACAAT	50.00	4.00	2.00		
in silico/in vitro	DPOL2	Forward primer	TTAGAGCTAAGGGCACCTGGA	52.38	6.00	1.00	270	*D.polymorpha*
		Reverse primer	TAAGGTTCAACCACCCCCGA	55.00	7.00	0.00		
in silico/in vitro	DPOL3	Forward primer	GAGCTAAGGGCACCTGGAAG	60.00	4.00	1.00	253	*D.polymorpha*
		Reverse primer	CCCCGAATCCTCCTTCCCTA	60.00	3.00	2.00		
in silico/in vitro	DPOL4	Forward primer	AGAGCTAAGGGCACCTGGAA	55.00	4.00	0.00	75	*D.polymorpha*
		Reverse primer	ACAAGCCCATGAGTGGTGAC	55.00	4.00	3.00		
Selected!	DROS1	Forward primer	GGAAACTGGTTGGTCCCGAT	55.00	4.00	2.00	188	*D.rostriformis bugensis*

		Reverse primer	GGCCCTGAATGCCCCATAAT	55.00	4.00	3.00		
in silico/in vitro	DROS2	Forward primer	TAGTGAGGGCGGATTTGGTG	55.00	2.00	0.00	185	*D.rostriformis bugensis* *Dreissena caputlacus* *Congeria kusceri* *Mytilopsis sallei*
		Reverse primer	AAAACTGATGACACCCGGCA	50.00	4.00	0.00		
in silico/in vitro	DROS3	Forward primer	TTAGTGAGGGCGGATTTGGT	50.00	2.00	0.00	82	*D. rostriformis bugensis* *Dreissena caputlacus*
		Reverse primer	TCTATGGCTGGCCCTGAATG	55.00	4.00	0.00		
in silico/in vitro	DROS4	Forward primer	TGTTCGGCGTTTAGTGAGGG	55.00	3.00	0.00	276	*D. rostriformis bugensis*
		Reverse primer	GGCACCGGCTAAAACAGGTA	55.00	4.00	2.00		
in silico/in vitro	DHAE6	Forward primer	TCTCCCTGTTTTAGCTGGCG	55	4	2	107	*Mycetophilidae sp* *Cordyla sp* *Thalamitoides sp*
		Reverse primer	AGTGCTGGTAAAGAATAGGATCTC	41.67	4	2		
in silico	DHAE7	Forward primer	CGGGTACAGGGTGAACTGTC	60	4	1	79	*Dikerogammarus villosus* *Dasyhelea sp.* *Ceratopogonidae sp.* *Mycetophilidae sp.* *Cordyla sp.*
		Reverse primer	AAAATAGCAAGATCAACAGAAGCA	33.33	4	0		
in silico	DHAE8	Forward primer	TGCTTCTGTTGATCTTGCTATTTT	33.33	4	0	103	*Dikerogammarus villosus* *Obesogammarus crassus*

		Reverse primer	TGGGGCGCGCATATTAAGAA	50	6	2		*Brevicornu fuscipenne* *Mycetophilidae sp.*
in silico	DHAE9	Forward primer	TGCTTCTGTTGATCTTGCTATTTTT	32	4	0	102	*Dikerogammarus villosus* *Pontogammarus obesus* *Obesogammarus crassus*
		Reverse primer	GGGGCGCGCATATTAAGAAC	55	6	1		
in silico	DHAE 10	Forward primer	ACCTCCTTCTTTAACTCTTCTTCTT	36	4	0	70	*Dikerogammarus villosus* *Cordyla sp.* *Mycetophilidae sp.*
		Reverse primer	GACAGTTCACCCTGTACCCG	60	4	2		
in silico/in vitro	DHAE 11	Forward primer	GGTCTGTCTTTATCACGGCCA	52.38	4	2	145	*Ochotona syrinx isolate*
		Reverse primer	AAGTGCTGGTAAAGAATAGGATCT	37.5	4	2		
in silico	DHAE 12	Forward primer	GGCAGGCCTGATATGGCTTT	55	8	2	78	*Dikerogammarus villosus*
		Reverse primer	AAGAAGAAGAGTTAAAGAAGGAGGT	36	4	0		
in silico/in vitro	DHAE 13	Forward primer	TCTCTCCCTGTTTTAGCTGGC	52.38	4	2	110	*Mycetophilidae sp.* *Cordyla sp.* *Thalamitoides sp.*
		Reverse primer	AAGTGCTGGTAAAGAATAGGATCTC	40	4	2		
in silico	DHAE 14	Forward primer	GGGTGAACTGTCTACCCTCC	60	7	3	72	*Dikerogammarus villosus* *Dasyhelea sp* *Ceratopogonidae sp.* *Mycetophilidae sp*
		Reverse primer	AAAAATAGCAAGATCAACAGAAGCA	32	4	0		

in silico/in vitro	DHAE 15	Forward primer	CCTGTTTTAGCTGGCGCTAT	50	5	3	105	*Mycetophilidae sp.* *Cordyla sp.* *Thalamitoides sp.*
		Reverse primer	TAAGTGCTGGTAAAGAATAGG ATCT	36	4	2		
in silico/in vitro	DVIL7	Forward primer	AGGCACTGGCTGAACAGTTT	50	5	3	72	*Cricotopus parbicinctus* *Phoridae sp* *Coelopa frigida* *Megaselia sp* *Sergentomyia barraudi*
		Reverse primer	GCAAGATCAACGGAAGCACC	55	4	0		
in silico	DVIL8	Forward primer	TGGCTGAACAGTTTACCCCC	55	7	3	111	*Gammarus cf. fossarum*
		Reverse primer	GCGCCGAGAATTGAAGAAGC	55	4	2		
in silico	DVIL9	Forward primer	TTGGCAGGTGCCATTACGAT	50	8	2	87	*Hirondellea gigas* Gammaridae *Harpactocrates ravastellus*
		Reverse primer	AAGAATAGGATCACCCCCGC	55	4	2		
in silico	DVIL10	Forward primer	TGTTAGGGAGGCCCGACATA	55	4	3	134	*Gammarus balcanicus* *Oecetis cinerascens*
		Reverse primer	GGGGTAAACTGTTCAGCCAGT	52.38	7	3		
in silico	DVIL11	Forward primer	GGCCCGACATAGCTTTTCCT	55	4	0	122	*Gammarus balcanicus* *Oecetis cinerascens*
		Reverse primer	GTAAACTGTTCAGCCAGTGCC	52.38	5	3		
in silico	DVIL12	Forward primer	AGGCCCGACATAGCTTTTCC	55	4	2	124	Ceratopogonidae *Strepsinoma*

		Reverse primer	GGTAAACTGTTCAGCCAGTGC	52.38	5	2		
in silico	DVIL13	Forward primer	GGAGGCCCGACATAGCTTTT	55	4	2	128	*Amphipoda sp* *Spintharus sp* *Metacrangonyx samanensis*
		Reverse primer	GGGGTAAACTGTTCAGCCAG	55	7	3		
in silico/in vitro	DVIL14	Forward primer	GGCACTGGCTGAACAGTTTAC	52.38	5	3	70	*Phoridae sp* *Simulium defoliarti*
		Reverse primer	CAAGATCAACGGAAGCACCG	55	4	2		
in silico	DVIL15	Forward primer	CTTGGCAGGTGCCATTACGA	55	8	0	85	*Amphipoda sp* *Reishia clavigera* *Thais clavigera*
		Reverse primer	AATAGGATCACCCCCGCCT	57.89	4	0		
in silico/in vitro	DVIL16	Forward primer	TCTTGGCAGGTGCCATTACG	55	8	2	86	Orthocladiinae
		Reverse primer	AATAGGATCACCCCCGCCTC	60	4	0		

Table S2.3: Tissue samples used in species specific primer testing.

Lane	Gel A	Gel B	Gel C	Gel D
1	*Dreissena rostriformis bugensis* tissue (5 ng/µl)	*Dreissena polymorpha* tissue (5ng/µl)	*Dikerogammarus villosus* tissue (3ng/µl)	*Dikerogammarus haemobaphes* tissue (5ng/µl)
2	Tissue dilution series 1:10 (~0.5 ng/µl)	Tissue dilution series 1:10 (~0.5 ng/µl)	Tissue dilution series 1:10 (~0.3 ng/µl)	Tissue dilution series 1:10 (~0.5 ng/µl)
3	Tissue dilution series 1:100 (~0.05 ng/µl)	Tissue dilution series 1:100 (~0.05 ng/µl)	Tissue dilution series 1:100 (~0.03 ng/µl)	Tissue dilution series 1:100 (~0.05 ng/µl)
4	Tissue dilution series 1:1000 (~0.005 ng/µl)	Tissue dilution series 1:1000 (~0.005 ng/µl)	Tissue dilution series 1:1000 (~0.003 ng/µl)	Tissue dilution series 1:1000 (~0.005 ng/µl)
5	*Anadonta anatina* (5 ng/µl)	*Anadonta anatina* (5 ng/µl)	*Crangonyx pseudogracalis* (5 ng/µl)	*Crangonyx pseudogracalis* (5 ng/µl)
6	*Sphaerium corneum* (5 ng/µl)	*Sphaerium corneum* (5 ng/µl)	Gammarus sp. (5 ng/µl)	Gammarus sp. (5 ng/µl)
7	*Dreissena polymorpha* (5 ng/µl)	*Dreissena rostriformis bugensis* (5 ng/µl)	*Dikerogammarus haemobaphes* (5 ng/µl)	*Dikerogammarus villosus* (5 ng/µl)
8	ddH_2O	ddH_2O	ddH_2O	ddH_2O

Table S2.4: Specimen biomass pre and post mesocosm experiment, *Dreissena rostriformis bugensis*

Tank number	Specimen number	Average specimen biomass prior to experiment (g)	Total specimen biomass before experiment (g)	Average specimen biomass after experiment (g)	Total specimen biomass after experiment (g)
Control	N/A	N/A	N/A	N/A	N/A
2	1	2	2	1.94	1.94
3	1	1.3	1.3	1.29	1.29
4	1	1.6	1.6	1.59	1.59
5	5	1.38	6.9	1.38	6.91
6	5	1.1	5.5	1.15	5.78
7	5	1.26	6.3	1.28	6.42
8	20	0.99	19.8	0.96	19.36
9	20	1.09	21.8	1.03	20.74
10	20	0.98	19.7	0.93	18.61

Table S2.5: Temperature measurements during mesocosm experiment, *Dreissena rostriformis bugensis*

Sample date	Sampling event	Room temp °C	Control tank Temp °C
12/1/2016	0 hours	15	9
12/1/2016	4 hours	14	9
12/1/2016	8 hours	14	8
13/1/2016	24 hours	14	9
18/1/2016	7 days	14	8
26/1/2016	15 days	15	9
1/2/2016	21 days	14	9
2/2/2016	22 days (24 hours after removal)	15	8
15/2/2016	28 days (7 days after removal)	14	9
22/2/2016	42 days (21 days after removal)	16	8

Table S2.6: Specimen biomass pre and post mesocosm experiment, *Dreissena polymorpha*

Tank number	Specimen number	Average specimen biomass prior to experiment (g)	Total specimen biomass before experiment (g)	Average specimen biomass after experiment (g)	Total specimen biomass after experiment (g)
Control	N/A	N/A	N/A	N/A	N/A
2	1	1.10	1.10	1.10	1.10
3	1	1.80	1.80	0.58	0.58
4	1	2.36	2.36	0.64	0.64
5	5	0.36	1.78	0.36	1.78
6	5	0.86	4.29	0.97	4.83
7	5	0.81	4.03	0.81	4.03
8	20	1.05	21.24	0.69	13.75
9	20	0.99	18.83	1.07	20.45
10	20	0.93	18.53	1.89	18.77

Table S2.7: Temperature measurements during mesocosm experiment, *Dreissena polymorpha*

Sample date	Sampling event	Room temp °C	Control tank Temp °C
12/7/2016	0 hours	15	8
12/7/2016	4 hours	14	8
12/7/2016	8 hours	15	8
13/7/2016	24 hours	15	9
17/7/2016	7 days	17	8
26/7/2016	15 days	16	8
1/8/2016	21 days	17	8
2/8/2016	22 days (24 hours after removal)	16	8
8/8/2016	28 days (7 days after removal)	14	9
22/8/2016	42 days (21 days after removal)	16	9

Table S2.8: Specimen biomass pre and post mesocosm experiment, *Dikerogammarus villosus*

Tank number	Specimen number	Total specimen biomass before experiment (g)	Total specimen biomass after experiment (g)
Control	N/A	N/A	N/A
2	1	0.0943	0.07
3	1	0.1799	0.9630
4	1	0.1864	0.0996
5	5	0.5097	0.3563
6	5	0.5627	0.3962
7	5	0.4647	0.4023
8	20	2.3291	0.9365
9	20	1.9828	0.5620
10	20	1.8622	0.6325

Table S2.9: Temperature measurements during mesocosm experiment, *Dikerogammarus villosus*

Sample date	Sampling event	Room temp °C	Control tank Temp °C
01/03/2016	0 hours	14	8
01/03/2016	4 hours	14	8
01/03/2016	8 hours	14	8
02/03/2016	24 hours	14	9
07/03/2016	7 days	15	8
15/03/2016	15 days	16	9
21/03/2016	21 days	14	8
22/03/2016	22 days (24 hours after removal)	14	9
29/03/2016	28 days (7 days after removal)	17	10
11/04/2016	42 days (21 days after removal)	16	9

Table S2.10: Specimen biomass pre and post mesocosm experiment, *Dikerogammarus haemobaphes*

Tank number	Specimen number	Total specimen biomass before experiment (g)	Total specimen biomass after experiment (g)
Control	N/A	N/A	N/A
2	1	0.8384	0.0864
3	1	0.0997	0.0390
4	1	0.1708	0.0345
5	5	0.8107	0.3273
6	5	0.6542	0.4068
7	5	0.6361	0.2152
8	20	2.3299	0.4719
9	20	2.4839	0.9191
10	20	2.3656	0.6535

Table S2.11: Temperature measurements during mesocosm experiment, *Dikerogammarus haemobaphes*

Sample date	Sampling event	Room temp °C	Control tank Temp °C
17/05/2016	0 hours	16	11
17/05/2016	4 hours	16	11
17/05/2016	8 hours	16	12
18/05/2016	24 hours	18	12
23/05/2016	7 days	16	12
31/05/2016	15 days	18	10
06/06/2016	21 days	15	12
07/06/2016	22 days (24 hours after removal)	16	14
13/06/2016	28 days (7 days after removal)	14	12
27/06/2016	42 days (21 days after removal)	16	14

Table S2.12: Species specific primer mismatches. Details of mismatch amplification with reference sequences taken from Genbank.

Primer Name	Sequence	Mis - matches	GenBank accession No.	Species
DV1_F	TCTTGGCAGGTCCCATTACG	1	KM208873	*D. villosus*
DV1_R	GAATCGGATCCCCCCCGCCT	2	EF570297	*D. villosus*
DH2_F	TAGGTCACGGGGGGGCTTCT	2	KM086443	*Thuridilla albopustulosa*
DH2_R	AAGTGTTGATAAAGAATAGG ATCA	3		
DH2_F	TTGGTCACAGTGGTGCTTCT	2	GU598226	*Hemicycla pouchadan*
DH2_R	AAATTCTGGTAAAGAATAGG ATCA	3		
DH2_F	TAGGTCACTGGAGTGCTTCT	2	KT794407	*Caucasotache a calligera*
DH2_R	AAATGCTGGTACAAAATAGG ATCT	3		
DP1_F	TAAAGCTAAGGGCACCTGGA A	1	AF510508	*D. polymorpha*
DP1_F	GAGAGCCAAGGGCACCTGGA A	2	JQ435817	*D. polymorpha*
DP1_R	AGCCCCTGAGTGGTGACAAT	1		

Table S2.13: *D. r. bugensis* mesocosm GLM. Null deviance for every model is 57.208 on 53.

Model	AIC	Estimate Std. Error | z value | Pr(>|z|)	Residual deviance P-val (1-pchisq)*	Residual deviance
Hours + Total biomass	53.495	0.001826 -1.811 0.0702 . 0.065035 2.111 0.0348 *	0.6136619	47.495 on 51
Hours + Density	53.609	0.001821 -1.808 0.0705 . 0.134796 0.065050 2.072 0.038 2 *	0.6091331	47.609 on 51
Total biomass	54.926	0.06278 2.054 0.040 *	0.5161544	50.926 on 52
Hours	58.17	0.001679 -1.726 0.084307	0.3916181	54.170 on 52
Density	55.027	0.06302 2.014 0.044 *	0.5121573	51.027 on 52

Table S2.14: *D. polymorpha* mesocosm GLM. Null deviance for every model is 41.654 on 53

Model	AIC	Estimate Std. Error | z value | Pr(>|z|)	Residual deviance P-val (1-pchisq)*	Residual deviance
Hours + Total biomass	32.827	0.002496 -1.842 0.0655 . 0.608773 1.761 0.0782 .	0.9978935	26.827 on 51
Hours + Density	32.823	0.002529 -1.841 0.0656 . 0.306233 1.969 0.0490 *	0.9978976	26.823 on 51
Total biomass	34.607	0.5625 1.702 0.0888 .	0.9921429	30.607 on 52
Hours	42.683	0.002088 -1.686 0.091790 .	0.9147678	38.683 on 52
Density	34.639	0.2849 1.880 0.0601 .	0.9920454	30.639 on 52

Table S2.15: *D. villosus* mesocosm GLM. Null deviance for every model is 68.744 on 53

Model	AIC	Estimate Std. Error | z value | Pr(>|z|)	Residual deviance P-val (1-pchisq)*	Residual deviance
Hours + Total biomass	31.115	0.002906 -0.408 0.682914 3.194821 0.718946 4.444 8.84e-06 ***	0.9991283	25.115 on 51
Hours + Density	32.082	0.002848 -0.402 0.688024 0.067877 4.549 5.39e-06 ***	0.998546	26.082 on 51
Total biomass	29.285	3.1680 0.7054 4.491 7.09e-06 ***	0.9993462	25.285 on 52
Hours	72.697	0.0015107 -0.216 0.829	0.06023596	68.697 on 52
Density	30.246	0.06665 4.596 4.31e-06 ***	0.9989001	26.246 on 52

Table S2.16: *D. haemobaphes* mesocosm GLM. Null deviance for every model is 74.563 on 53

Model	AIC	Estimate Std. Error z value Pr(>|z|)	Residual deviance P-val (1-pchisq)*	Residual deviance
Hours + Total biomass	65.505	0.001663 -0.839 0.40163 0.411376 3.254 0.00114 **	0.1936306	59.505 on 51
Hours + Density	61.099	0.001757 -0.875 0.381665 0.053151 3.418 0.000631 ***	0.322359	55.099 on 51
Total biomass	64.222	0.4065 3.245 0.00118 **	0.202736	60.222 on 52
Hours	78.024	0.001424 -0.732 0.464	0.02404275	74.024 on 52
Density	59.887	0.05243 3.407 0.000658 ***	0.3310254	55.887 on 52

2.8 Supplementary Information II

DNA Extraction method – Protocol for DNA extraction from filter papers modified from Brolaski et al., (2008).

Lysis solution 1 – 0.12M guanidine thiocyanate, 0.181 M trisodium phosphate

Lysis solution 2 – 5 M sodium chloride, 0.5 M Tris base, 4% SDS

Precipitation solution – 5 M ammonium acetate, 0.12 M alluminium ammonium sulphate dodecahydrate

Binding solution – 5 M guanidine HCl, 0.03 M Tris HCl, 9% isoproponol

Wash solution – 0.01 M Tris HCl, 0.5 M sodium Chloride, 75% ethanol

Elution Buffer – 1X TE buffer (1:10)

1g 30mesh garnet beads, 1g fine sand into 7ml tube

Add filter paper

925 µl Lysis solution 1 and 75 µl Lysis solution 2

Qiagen Tissue lyser 5 minutes, 30 bps

Centrifuge 4000g, 1min

Pipette off supernatant into clean 2ml tube

Add 250 µl Precipitation solution, vortex

Chill on ice for 5 mins

Centrifuge 10000g, 1min

Pipette off supernatant into clean 2ml tube

Add x1.5 volume of Binding solution, vortex

Pipette 650 µl into spin column, centrifuge 10000g, 1 min, discard flow through

Repeat step 12 until all solution has gone through spin column

Add 500 µl of Wash solution, centrifuge 10000g, 1 min, discard flow through

Centrifuge spin column 10000g, 2 min

Place spin column in fresh collection tube

Add 100 µl of Elution buffer (TE, ddH_2O) leave for 5 minutes

Centrifuge 10000g, 1 min.

2.9 Supplementary Information III

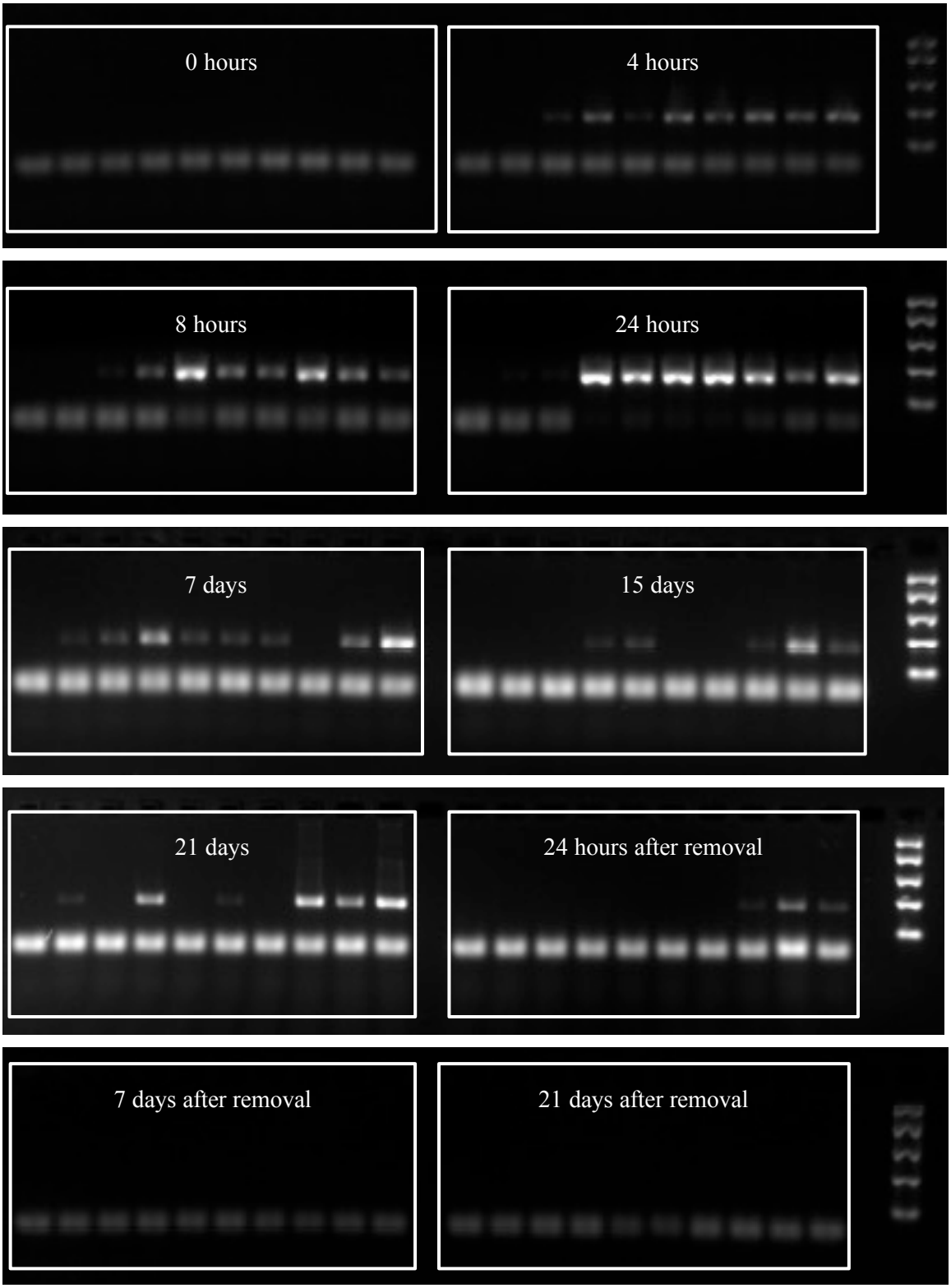


Figure S 2.1: Agarose gel images from *Dreissena rostriformis bugensis* mescosom sampling. Lane 1 – control tank, Lane 2-4 = 1 specimens of INNS, Lanes 5-7 = 5 specimens and Lanes 8 – 10 = 20 specimens. The final lane is DNA EasyLadder I (Bioline, UK) with fragment sizes of 100 bp, 250 bp, 500 bp, 1000 bp and 2000 bp.

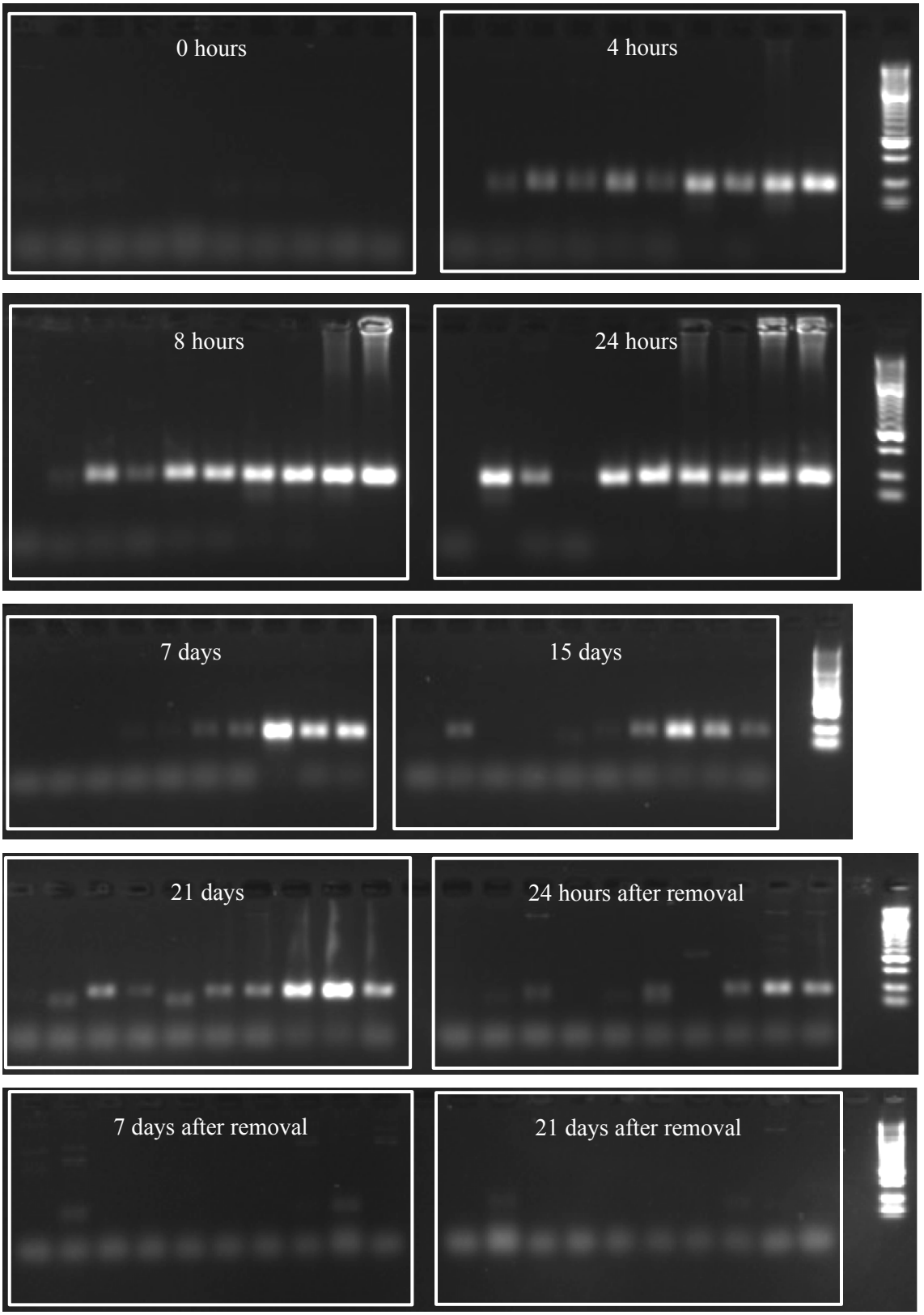


Figure S 2.2: Agarose gel images from *Dreissena polymorpha* mescosom sampling. Lane 1 – control tank, Lane 2-4 = 1 specimens of INNS, Lanes 5-7 = 5 specimens and Lanes 8 – 10 = 20 specimens. The final lane is DNA Hyperladder 50bp (Bioline, UK) with fragment sizes of 50 bp, 100 bp, 200 bp, 300 bp - 2000 bp.

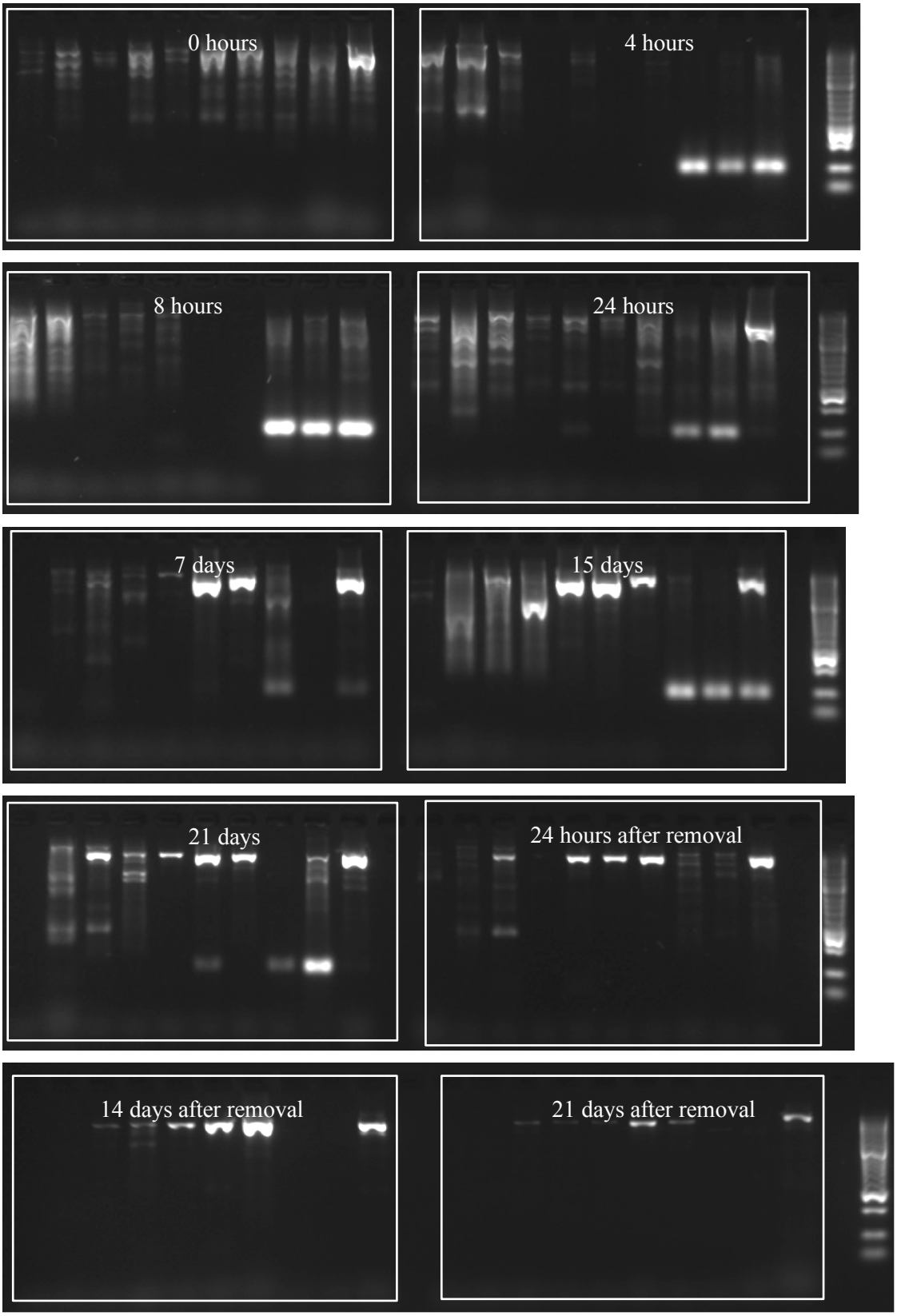


Figure S 2.3: Agarose gel images from *Dikerogammarus villosus* mescosom sampling. Lane 1 – control tank, Lane 2-4 = 1 specimens of INNS, Lanes 5-7 = 5 specimens and Lanes 8 – 10 = 20 specimens. The final lane is DNA Hyperladder 50bp (Bioline, UK) with fragment sizes of 50 bp, 100 bp, 200 bp, 300 bp - 2000 bp.

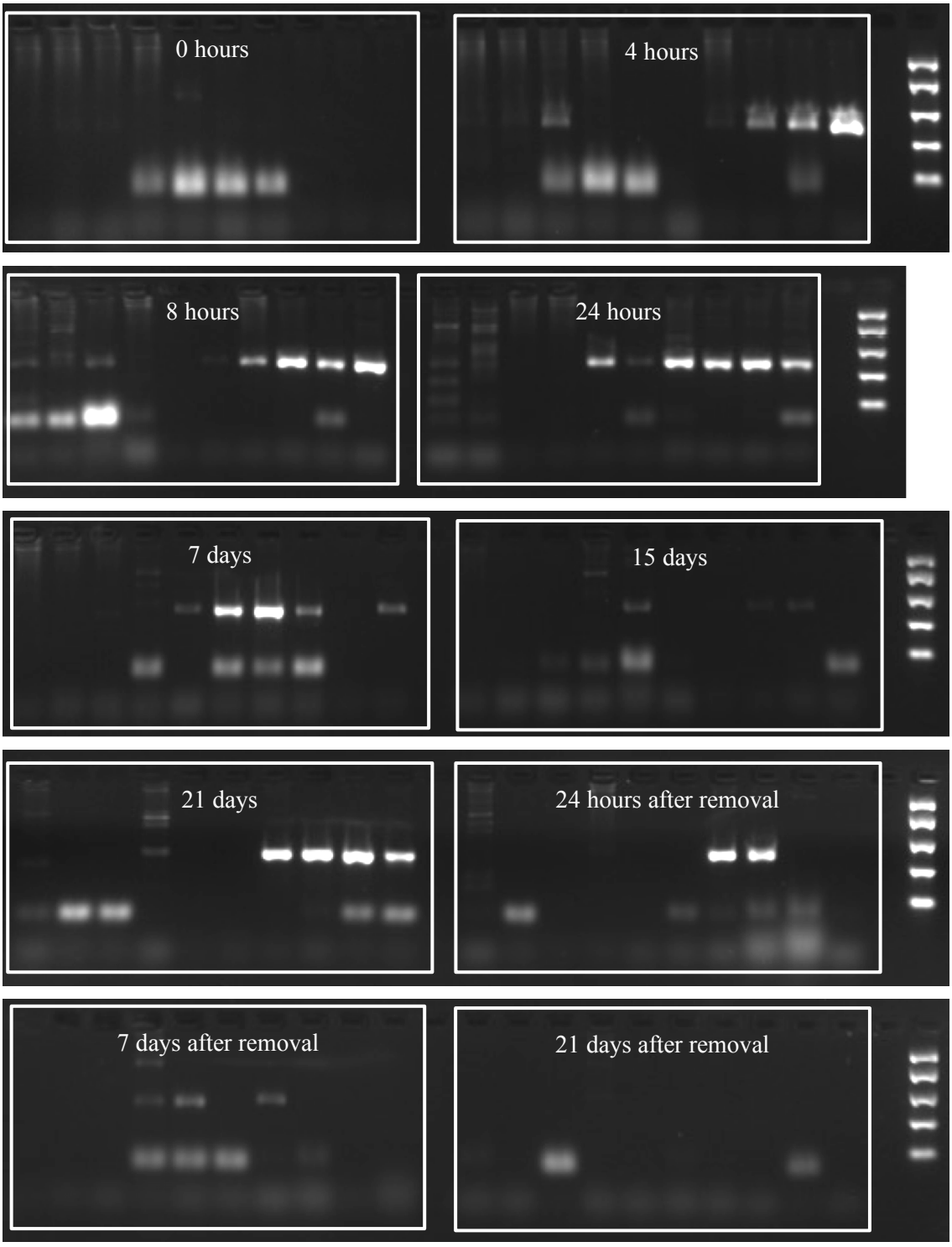


Figure S 2.4: Agarose gel images from *Dikerogammarus haemobaphes* mescosom sampling. Lane 1 – control tank, Lane 2-4 = 1 specimens of INNS, Lanes 5-7 = 5 specimens and Lanes 8 – 10 = 20 specimens. The final lane is DNA EasyLadder I (Bioline, UK) with fragment sizes of 100 bp, 250 bp, 500 bp, 1000 bp and 2000 bp.

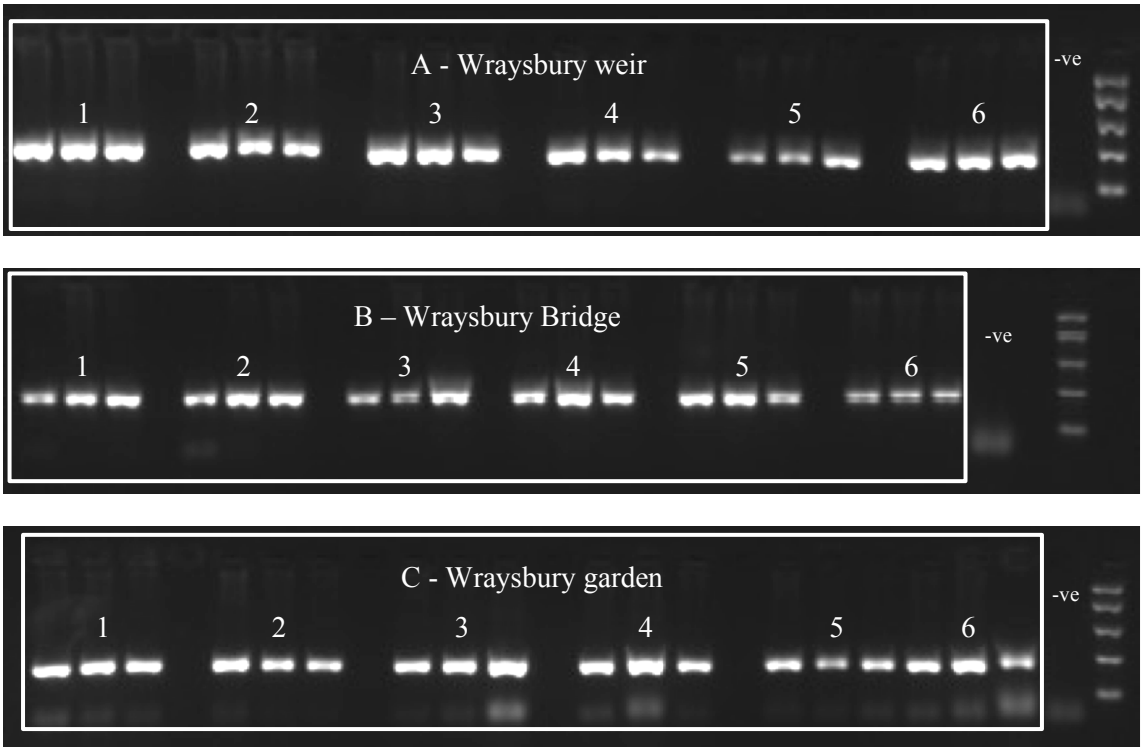


Figure S 2.5: Agarose gel images from *Dreissena rostriformis bugensis* field samplescollected from the River Wraysbury. A – Wraysbury Weir, B – Environment Agency kicknet sampling site and C – Wraysbury Gardens. The final lane is DNA EasyLadder I (Bioline, UK) with fragment sizes of 100 bp, 250 bp, 500 bp, 1000 bp and 2000 bp.

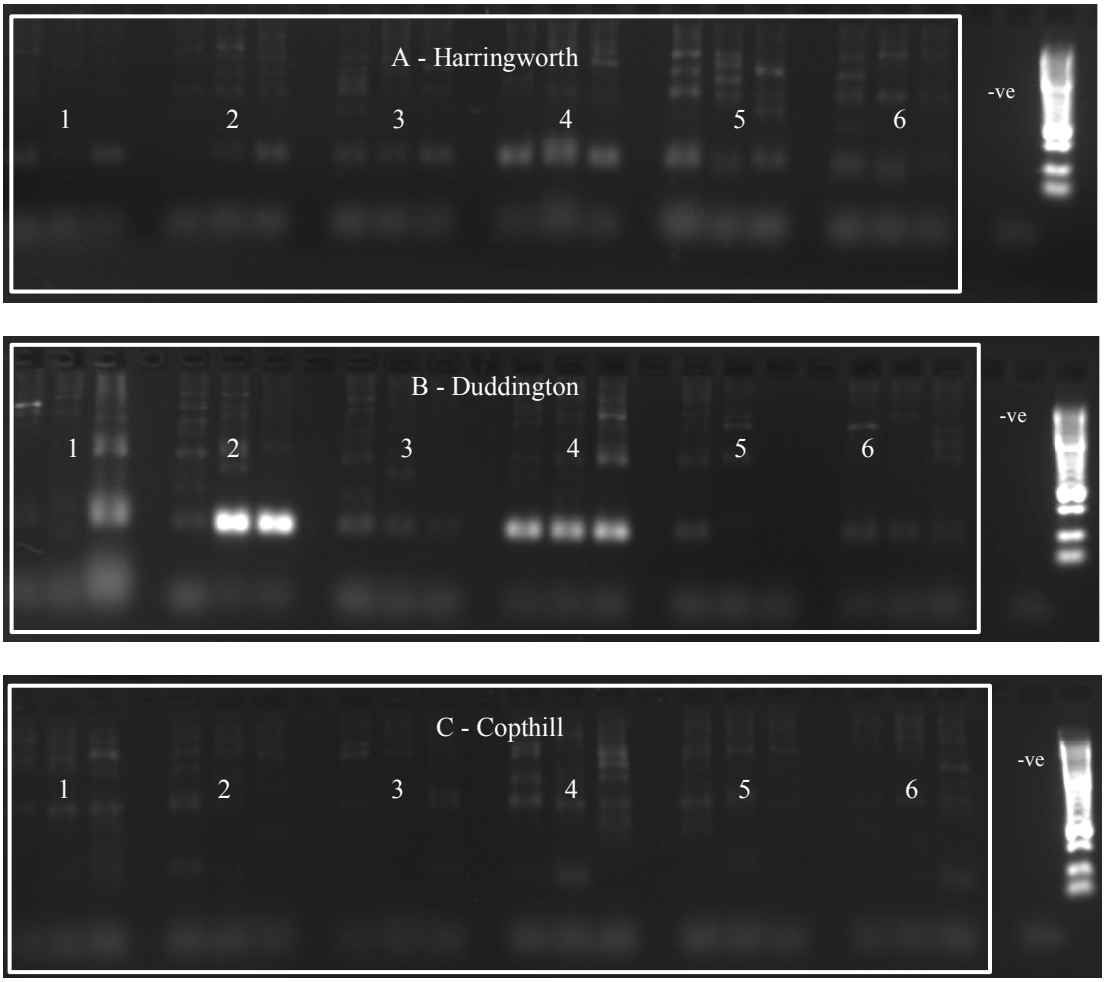
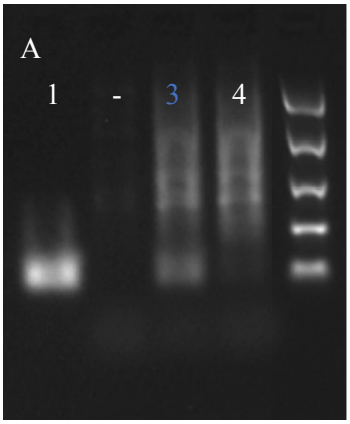


Figure S 2.6: Agarose gel images from *Dreissena polymorpha* field samplescollected from the River Welland. A – Harringworth, B – Duddington Mill and C – Copthill Bridge. The final lane is DNA Hyperladder 50bp (Bioline, UK) with fragment sizes of 50 bp, 100 bp, 200 bp, 300 bp - 2000 bp.



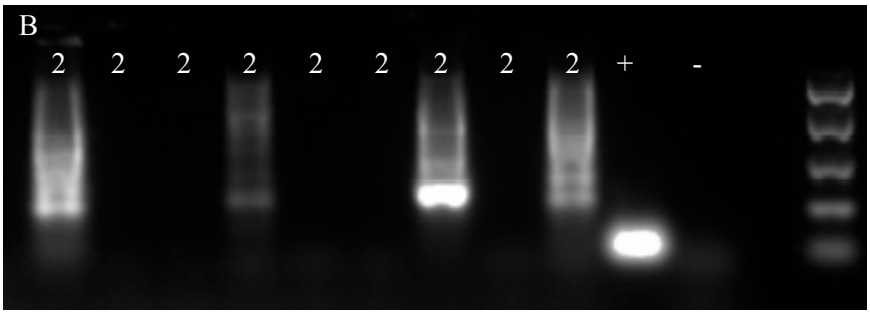


Figure S 2.7A and S 2.7B: Agarose gel images from *Dikergammarus villosus* field samples collected from 1 – *D. villosus* tissue sample, 2 - Grafham Water, 3 - Pitsford Water and 4 - Wroxham Broad. Note S7B includes 3 replicated at each of the 3 sites sampled at Grafham Water. The final lane is DNA EasyLadder I (Bioline, UK) with fragment sizes of 100 bp, 250 bp, 500 bp, 1000 bp and 2000 bp.

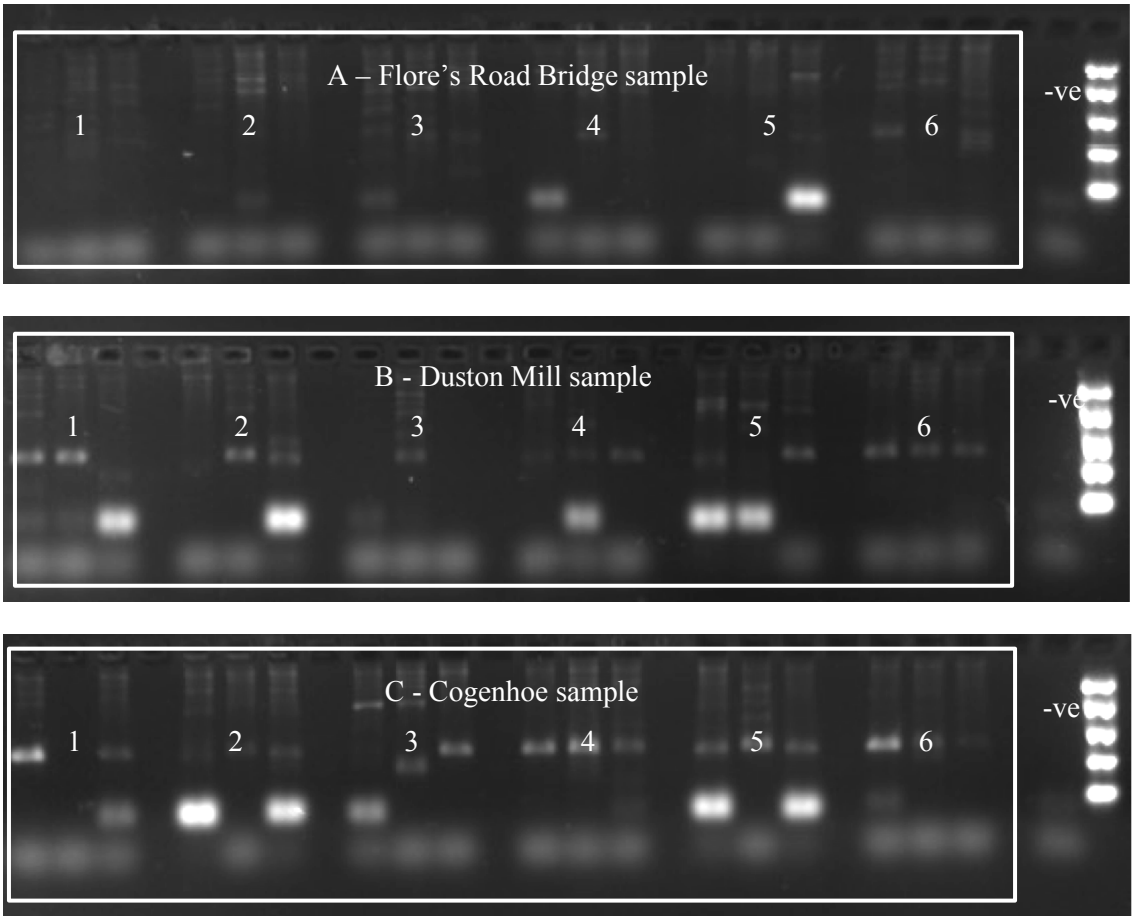


Figure S 2.8: Agarose gel images from *Dikergammarus haemobaphes* field samplescollected from the River Nene. A – Flore's Road Bridge, B – Duston Mill and C – Cogenhoe Mill. The final lane is DNA EasyLadder I (Bioline, UK) with fragment sizes of 100 bp, 250 bp, 500 bp, 1000 bp and 2000 bp.

Chapter 3 What is the optimal approach for monitoring aquatic invasive non-native species? A comparison of established and environmental DNA methods for monitoring the quagga mussel

3.1 Abstract

Rapid increase in trade and transport has caused a surge in the number of successful invasions by invasive non-native species (INNS) to new environments. Traditional monitoring for new INNS relies on their observation or collection, which is often only successful after species become established. Environmental DNA (eDNA) approaches, which detect free-floating DNA shed by organisms into their environment, are highly promising for sensitive detection of INNS during the early stages of invasion but few studies have directly compared the sensitivity of different strategies.

Environmental DNA can be used to target a single species ("targeted detection") or a whole community ("passive detection" methods such as metabarcoding). In this study, we compare these molecular approaches with traditional monitoring methods for the detection of quagga mussel, *Dreissena rostriformis bugensis*. Following mesocosm validation tests for a new qPCR assay, PCR, metabarcoding and kick-net sampling were directly compared in the field along a density gradient from the main source population to 2.75 km downstream in the River Wraysbury.

All three molecular tools outperformed traditional monitoring for quagga mussel detection in the field. PCR and qPCR both had 100% detection rate in all samples, and outperformed metabarcoding for detecting the target species at low densities. For qPCR and metabarcoding there was a strong, significant negative relationship between distance from source population, and DNA copy number / read count respectively.

We conclude that all three molecular methods are more sensitive than traditional kick sampling for detection of *D. r. bugensis* in flowing water, and both qPCR and metabarcoding are suitable for obtaining estimates of relative abundance. Targeted methods had a greater sensitivity than metabarcoding, but metabarcoding has the advantage of providing information on non-target taxa, which could be invaluable for detection of additional INNS and investigating impacts on communities.

Note:

This chapter has been written for submission to Molecular Ecology and Evolution, and so the abstract is written in the required journal format.

3.2 Introduction

Invasive non-native species (INNS) have a range of damaging effects on ecosystems outside their native range. Their effects range from outcompeting native species to causing damage to infrastructure (e.g. Dressenid mussel biofouling) (Roy et al., 2014). In the UK, up to £43.5 million is spent per annum on controlling freshwater INNS alone (Oreska and Aldridge, 2003). Early detection of an INNS is crucial, as it is during this stage that eradication or containment methods are most effective (Hulme, 2006) and cost effective (Roy et al., 2014). Current methods of monitoring freshwater macroinvertebrates rely on specimen collection and correct morphological identification, which are not always suitable to species in low abundance. This is particularly relevant to new introductions of INNS, which are often unnoticed until they become established (Hänfling et al., 2011) or cryptic to identify and may be misidentified as native species (Blackman et al., 2017a). Alternative methods that offer rapid, accurate and cost-effective detection of new INNS need to be explored.

The use of molecular tools to identify DNA taken from environmental samples (i.e. environmental DNA or "eDNA") has been a major research focus in biodiversity monitoring over the last 10 years (Ficetola et al., 2008, Thomsen et al., 2012b). Two broad approaches can be taken for detection of species of interest. Firstly a "targeted" approach utilises species specific primers to detect a single species, for example via standard PCR, quantitative PCR (qPCR) or droplet digital PCR (ddPCR). Secondly, a "passive approach" through metabarcoding. Metabarcoding utilises widely conserved primers to amplify DNA from a group of taxa, for example fish and other vertebrates (e.g. Kelly et al., 2014; Miya et al., 2015; Simmons et al., 2015; Hänfling et al., 2016; Valentini et al., 2016), zooplankton (e.g. Brown et al., 2016), insects (e.g. Yu et al., 2012., Schneider et al., 2016) or macroinvertebrates (e.g. Elbrecht et al., 2016., Blackman et al., 2017a., see e.g. Lawson Handley., 2015; Creer et al., 2016; Valentini et al., 2016; Deiner et al., 2017 for reviews). Amplicons are then sequenced using High Throughput Sequencing (HTS) platforms and bioinformatically assigned to samples and taxa. Metabarcoding provides the opportunity for detecting multiple species at the same time including those which have not been listed as INNS or have not been prioritised for monitoring (Simmons et al., 2015; Blackman et al., 2017a; Deiner et al., 2017). Comparison between traditional methods for biodiversity monitoring and both targeted (e.g. Dejean et al., 2012; Tréguier et al., 2014) and passive (e.g. Hänfling et al., 2016; Smart et al., 2015; Wilcox et al., 2016) molecular approaches have demonstrated an increased probability of detection using eDNA. However, to our knowledge, no previous studies have

compared the sensitivity of traditional methods and both targeted and passive molecular approaches for the detection of invasive non-native species.

The targeted approach was first successfully used by Ficetola et al., (2008) to detect invasive American bull frogs, *Lithobates catesbeiana*, in lentic waterbodies and since has been successfully applied to over 100 target taxa (Blackman et al., 2017b *in review*). The majority of targeted studies have used either standard or qPCR for detection, and there are a number of pros and cons associated with both method. Standard PCR (PCR) is a well-established and widely practiced method for detecting presence/absence of target DNA by visualisation of PCR products on agarose gels. qPCR has on many occasions be shown to be more sensitive than PCR, (Nathan et al., 2014; Simmons et al., 2015; De Ventura., et al 2016) so may be more suitable for detecting species at low density, as in the early stages of invasion. qPCR has the additional advantage of providing quantitative information on the number of DNA copies in a sample, and a number of eDNA studies have demonstrated significant correlations between DNA copy number and biomass or abundance (Amphibians, Thomsen et al., 2012b; Fish, Takahara et al., 2012; Gastropods, Goldberg et al., 2013). However, qPCR requires the construction of calibration curves from known standard concentrations and is substantially more expensive and time-consuming than standard PCR (Nathan et al., 2014). Direct comparisons of standard and qPCR for targeted detection of species from eDNA have revealed drawbacks of qPCR including high risk of errors from standard curve construction, poor reproducibility between laboratories (Nathan et al., 2014), higher risk of false positives due to increased sensitivity to contamination (De Ventura et al., 2017), and higher sensitivity to PCR inhibitors (Davy et al., 2015) can be problematic in qPCR, especially when dealing with highly sensitive samples such as eDNA. In some cases, therefore, PCR may be preferable to qPCR for targeted detection as it can be more robust and less prone to false positives or false negatives than qPCR (De Ventura et al., 2017; Davy et al., 2015).

The potential for eDNA metabarcoding for INNS monitoring has been demonstrated in a small number of studies to date (e.g. Simmons et al., 2015; Brown et al., 2016; Cannon et al., 2016; Blackman et al., 2017a; Klymus et al., 2017; Borrell et al., 2017). For example, studies have shown presence of new INNS in the wild, which had been previously overlooked by traditional methods or not targeted by molecular methods. In a river macroinvertebrate metabarcoding study, Blackman et al., (2017a) detected a non-native Gammaridae species, *Gammarus fossarum*, which had been missed by traditional methods for over 50 years in the UK and is now widespread and

common. Similarly, Simmons et al., (2015) detected Northern Snakeshead, *Channa argus*, in the Muskingum River Watershed, Ohio, which unlike Asian Carp had not been included on the INNS priority lists. Metabarcoding also has the potential to act as a surveillance tool and has already been applied successfully by some studies to monitor high risk pathways, such as ballast (Ardura et al., 2015b; Zaiko et al., 2015), and the live bait trade (Mahon et al., 2014) and locations such as ports (Brown et al., 2016; Borrell et al., 2017). Although a few studies have directly compared the performance of targeted approaches for detection of INNS or other focal species (e.g. Nathan et al., 2014; Davy et al., 2015; De Ventura et al., 2017), to our knowledge, only one study has compared targeted and passive approaches for INNS monitoring (Simmons et al., 2015). As well as detecting previously unrecorded *C. argus* as mentioned above, this study demonstrated the higher sensitivity of ddPCR compared to PCR or metabarcoding for detection of bighead carp, *Hypopthalmicthys nobilis* (Simmons et al., 2015). Further comparison of the methods is needed to evaluate whether sensitivity is comparable and to better inform management strategies.

The quagga mussel, *Dreissena rostriformis bugensis*, is a highly invasive bivalve from the Ponto-caspian region. It has been described as an 'ecosystem engineer', having a significant impact on all trophic levels within the environments it has invaded (Karatayev et al., 2002; Karatayev et al., 2007; Roy et al., 2014). *Dreissena rostriformis bugensis* also scored the highest in a recent horizon scanning exercise for potential invasions to the UK in terms of impact, risk of invading and establishing (Roy et al., 2014), and was detected later the same year in the Wraysbury Reservoir. Due to its wider tolerance of environmental variables, *D. r. bugensis* has the potential to spread rapidly, and occupy an even greater range of habitat than closely related zebra mussels, *Dreissena polymorpha*, which are already widespread in the UK (Quinn et al., 2014; Nalepa et al., 2010; Gallardo and Aldridge 2013). Quagga mussels are therefore at the very top of the UK priority list for early detection and monitoring of spread. Motivated by a need for cost effective tools for quagga mussel detection, we previously designed a standard PCR assay, which provided 100% probability of detection in field trials (Blackman et al., 2017c *in review*). Here we further developed this assay for abundance estimation using qPCR and compare the sensitivity of both targeted methods to metabarcoding and traditional kick-net sampling. Although other recent studies have developed PCR and qPCR assays for quagga mussel (De Ventura et al., 2017), to our knowledge the current study is the first to fully evaluate the sensitivity of both passive and targeted as well as traditional monitoring approaches for quagga mussel detection and abundance estimation. Specifically, we first developed and validated a dye-based qPCR assay for quagga mussels in mesocosm experiments and field trials. Secondly, we directly compared the three molecular methods (PCR, qPCR and metabarcoding) and traditional kick-net sampling for

detection of quagga mussels in the field. Finally, we investigated whether qPCR and metabarcoding are suitable for estimating relative abundance of quagga mussels by testing for a relationship between DNA copy number (qPCR) or read count (metabarcoding) and distance along a density gradient from the main source population to 2.75 km downstream in the River Wraysbury, Berkshire, UK. We hypothesized that (1) probability of detection will be higher for eDNA methods than traditional methods, (2) targeted methods will have higher probability of detection than the passive method, (3) qPCR will be more sensitive than standard PCR and (4) both read count and DNA copy number will decline with distance from the main source population.

3.3 Methods

3.3.1 Sample collection, DNA capture and extraction

Quagga mussels were first detected in the Wraysbury Reservoir, Berkshire in October 2014. The reservoir is considered to be the main source population, and mussels are spreading along the River Wraysbury, which is a tributary of the River Colne in the Thames catchment. Sampling was carried out in April 2016, working in an upstream direction to ensure DNA availability was not altered by previous kick-net sampling higher in the catchment. Six 1.5 L water samples and six kick-net samples were obtained from each of three main sampling sites along the River Wraysbury: Wraysbury Weir (WW, 0.61 km downstream of Wraysbury Reservoir), Wraysbury Bridge (WB, 1.10 km), and Wraysbury Gardens (WG, 2.75 km,) (Fig. 3.1) in order to directly compare established and molecular methods. An additional four 1.5 L samples were collected for qPCR validation and analysis (Fig. 3.1) from the following locations: Reservoir Outfall (RO, 0.2 km downstream from Wraysbury Reservoir), Moor Lane (ML, 1.7 km), Hale Street (HS, 2.72 km), and Upstream Thames confluence (UT, 2.75 km, see Fig. 3.1). An upstream sample from above the Wraysbury Reservoir outfall was also collected to act as a negative control. Sample bottles filled with ddH_2O were also taken into the field and later filtered as sample blanks. Kick-sampling was performed using the standard method for macroinvertebrate community monitoring, which is a three minute kick sample followed by a one minute hand search (Murray-Bligh, 1999). Kick samples were placed in labelled 42 fluid oz. sterile Whirl-pak® bag (Cole-Palmer, Hanwell, London) and stored in a cool box before being frozen at the laboratory. Each sample was defrosted and washed with purified water prior to analysis, then sorted for *D. r. bugensis* by eye. All specimens of *D. r. bugensis* were counted and subsequently stored in 100% ethanol. Each water sample was collected by pooling 3 x 500 ml collected from the surface across

the river width, into sterile sampling bottles. 500 ml of each sample was then filtered and extracted separately to allow replication within each eDNA sampling event (n = 66). Samples were vacuum filtered through sterile 47 mm diameter 0.45 µm cellulose nitrate membrane filters with pads (Whatman, GE Healthcare, UK) within 24 hours of collection, using Nalgene filtration units (Thermo Fisher Scientific) in combination with a vacuum pump (15~20 in. Hg, Pall Corporation) in a dedicated eDNA laboratory at the University of Hull, UK. All sampling and filtration equipment was sterilized in 10% commercial bleach solution for 10 minutes then rinsed with 10% MicroSol detergent (Anachem, UK) and purified water between samples. DNA extractions were carried out using PowerWater® DNA Isolation Kit (MoBio Laboratories, Inc. Carlsbad, USA) following the manufacturer's instructions.

3.3.2 Targeted Assays: standard PCR

In a previous study, we developed and validated a standard PCR assay for *D. r. bugensis* that targets a 188 bp fragment of the mitochondrial COI region using primer pair DRB1 (DRB1_F - GGAAACTGGTTGGTCCCGAT and DRB1_R GGCCCTGAATGCCCCATAAT, (Blackman et al., 2017c, *in review*). This assay is highly sensitive for PCR (Limit of Detection (LoD) 0.0003 ng/µl) and does not cross amplify closely related Driessenid or native mussel species (Blackman et al 2017c, *in review*). All samples (n = 66), sampling blanks (n = 6), filter blanks (n = 6), and PCR blanks (n = 6) were tested with DRB1, including DNA from the closely related zebra mussel, *Dreissena polymorpha* acting as an additional negative control and to check for cross contamination. PCRs were carried out in triplicate in 25 µl volumes with MyTaq Red Mix Taq (Bioline, UK) containing: 10 µM of each primer, and 2 µl of undiluted DNA template. PCRs were performed on an Applied Biosystems Veriti Thermal Cycler with the following profile: initial denaturation at 94°C for 3 minutes, followed by 37 cycles of denaturation at 94°C for 30s, annealing at 65°C for 1 minutes and extension at 72°C for 1 minute 30s, with a final extension time of 10 minutes at 72°C. Detection was then confirmed by running PCR products on a 1.5% agarose gel stained with GelRed (Cambridge Biosciences, UK). For successful detection, we set criteria that at least 2 of the 3 replicates must amplify, and any samples which did not meet this requirement were re-tested.

3.3.3 Targeted Assays: qPCR assay development and validation in mesocosms

The DRB1 assay was further developed using fluorescent dye-based qPCR. We carried out a serial dilution of DNA from tissue samples (from 1 to 1 x 10^{-7} ng/µl) and a mesocosm experiment

to ensure the assay worked for the detection of *D. r. bugensis* DNA only with no cross amplification. In brief, the mesocosm experiment monitored DNA production from mussels in three densities (1, 5 and 20) over 21 days, with 3 replicates of each density and a control tank with no mussels present (See Chapter 2 section 2.3 for further details). Fifteen litre tanks were sampled for 21 days with the specimens present at: 0hrs, 4hrs, 8hrs, 24hrs, 48 hrs, 72 hrs, 7 days, 15 days and 21 days. On day 21 the specimens were removed and sampling continued at: 22 days, 28 days, 35 days, 42 days (n = 130). Two hundred millilitres of water was sampled at each time point and a single filtration was carried out for each tank sample using the same protocol as the field samples. Extractions for these samples were carried using a protocol modified from Bolaski et al., (2008) (for the full extraction protocol, see Supplementary Information I).

Quantitative PCR reactions were carried out in triplicate on a StepOne-PlusTM Real-Time PCR machine (Applied Biosystems) in 25 µl reaction volumes. Primers (DRB1) and reagent concentration were the same as for the PCR reaction, however for qPCR, MyTaq was replaced with Power Up SYBR Master Mix (Fisher Scientific, UK). qPCRs were performed with the following profile: 2 minutes at 50°C, 10 minutes at 95°C, and 50 cycles of 15 s at 95°C and 1 minute at 60°C followed by melt curve analysis. A 300 bp laboratory synthesized fragment of the *D. r. bugensis* COI gene was designed using gBlocks® (Integrated DNA Technologies Inc.) to be used as the standards. DNA copy number for gBlock fragments were estimates by using Avogadro's number (number of copies = (500 ng of DNA × 6.022 × 1023) / (300 bp fragment × 1 × 109 × 650). This formula assumes that the average weight of a base pair is 650 Da. The lypholised gBlock was re-suspended in 50 µl of sterile 1x TE buffer, which made a stock concentration of 5.0 x 10^{10} copies. From this stock solution, 1 five-fold dilution followed by eight 10-fold dilutions were made to produce the standard curve (10 to 10^{8} copies). All samples (n = 66), sampling blanks (n = 6), filter blanks (n = 6), and PCR blanks (n = 6) were tested with DRB1 qPCR assay, including DNA from the closely related zebra mussel, *Dreissena polymorpha* on each PCR plate acting as an additional negative control and to check for cross contamination. Amplification curves, Cq values and melting curves were analysed using StepOne-PlusTM software. DNA copy number for each sample was calculated using Cq number and standard curve results (See supplementary Information II for full MIQE checklist).

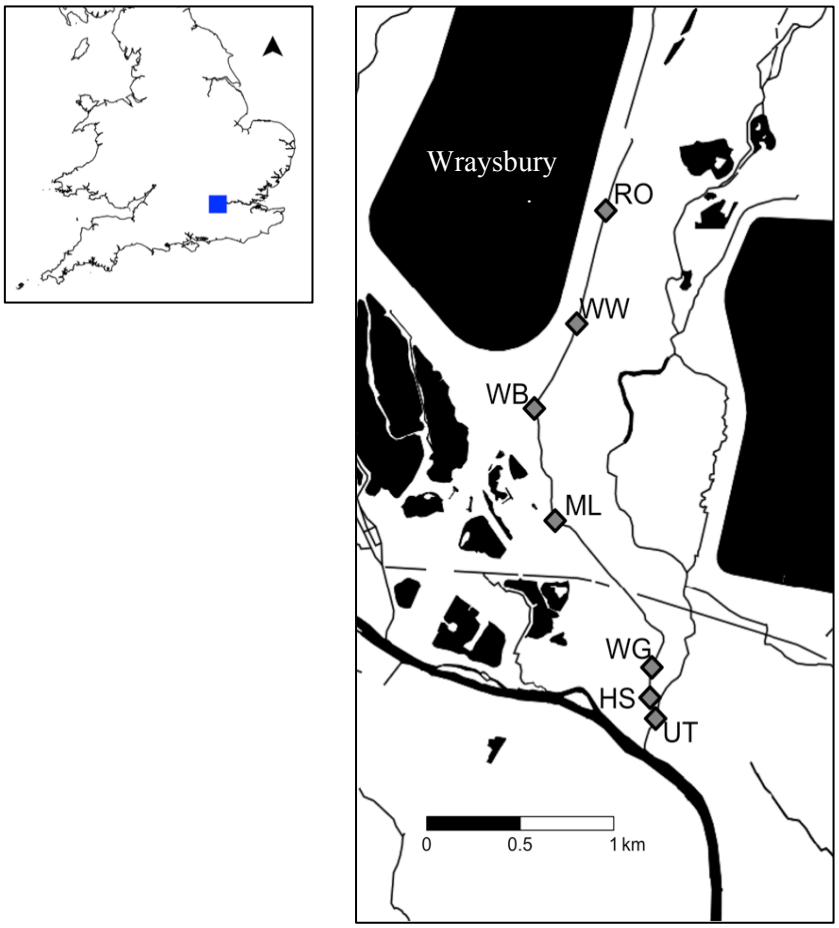


Figure 3.1: Sampling sites on the Wraysbury River, downstream of the main source population of *D. r. bugensis* in the Wraysbury Reservoir. Sampling sites were the Reservoir Outfall (RO, 0.2 km downstream of the reservoir), Wraysbury Weir (WW, 0.61 km), Wraysbury Bridge (WB, 1.10 km), Moor Lane (ML, 1.7 km), Wraysbury Gardens (WG, 2.7 km), Hale Street (HS, 2.72 km), and Upstream Thames confluence (UT, 2.75 km). Full grid references are given in Table S3.1.

3.3.4 Passive detection: eDNA metabarcoding

The COI region was chosen for metabarcoding as it has the broadest taxonomic coverage for macroinvertebrates in public sequence databases and is the most widely used DNA barcode for taxonomic discrimination in this group. A 313 bp fragment was targeted using the primers described in (Leray et al., 2013). A two-step protocol was used for library preparation, as outlined below.

First step PCRs were performed with the primers jgHCO2198: TAIACYTCIGGRTGICC RAARAAYCA and mICOIintF: GGWACWGGWTGAACWGTWTAYCCYCC (Leray et al.,

2013). PCRs were carried out in 50 µl volumes containing 12.5 µl Taq DNA polymerase (Sigma Aldrich, UK), 0.2 µM of each primer, 1.5 mM $MgCl_2$, 2 µM dNTPs, and 2 µl of undiluted DNA template. PCRs were performed on an Applied Biosystems Veriti Thermal Cycler machine with the following profile: initial denaturation 95°C for 2 mins, followed by 35 cycles of denaturation 95°C for 15s, annealing at 50°C for 30s and extension at 72°C for 30s, with a final extension time of 10 min at 72°C. PCR and filtering blanks (n=6 of each) and positive controls of DNA from species that are terrestrial (Osmia bicornis, Linnaeus, 1758, n=2) or not found in the sampled water bodies (Triops cancriformis, Bosc, 1801, n=2) were included in all PCRs. PCR products were confirmed by gel electrophoresis on a 1.5% agarose gel stained with GelRed (Cambridge Bioscience, UK). Post first PCR clean-up was carried out using ZR-96 DNA Clean-up Kit™ (Zymo Research, USA) following the manufacturer's protocol.

Individual Nextera tags (Illumina, UK) were then added in a second PCR with 10 µM of each tagging primer and 1 µl of purified PCR product. PCR settings were: initial denaturation 95°C for 2 min, followed by 8 cycles of denaturation 95°C for 15s, annealing at 55°C for 30 secs and extension at 72°C for 30 secs, with a final extension time of 10 mins at 72°C. Samples were then normalised using SequalPrep Normalization plates (Invitrogen, UK). Each plate of samples was then pooled and the library made with equimolar concentrations of each plate. The final pooled library was concentrated and then cleaned using the QIAquick Gel Extraction Kit (Qiagen, UK) following the manufacturer's protocol. The library was run at 13 pM concentration with 20% PhiX, using the Illumina 2 x 250bp V2 chemistry, on an Illumina MiSeq at the in-house facility at the University of Hull.

Processing of Illumina read data and taxonomic assignment were performed using a custom bioinformatics pipeline MetaBEAT (metaBarcoding and eDNA Analysis Tool) v 0.8 (https://github.com/HullUni-bioinformatics/metaBEAT) described previously (Hänfling et al., 2016, Blackman et al., 2017a), with minor modifications. The program Trimmomatic 0.32 (Bolger et al., 2014) was used for quality trimming and removal of adapter sequences from the raw Illumina reads. Average read quality was assessed in 5 bp sliding windows starting from the 3'-end of the read and reads were clipped until the average quality per window was above phred 30. All reads shorter than a defined minimum read length (313bp) were discarded. Sequence pairs were subsequently merged into single high-quality reads using the program FLASH 1.2.11 (Magoč & Salzberg., 2011). The remaining reads were screened for chimeric sequences against a reference database using the 'uchime_ref' function implemented in vsearch 1.1. To remove

redundancy, sequences were clustered at 100% identity using vsearch 1.1. Clusters represented by less than 5 sequences were considered sequencing error and were omitted from further analyses. Non-redundant sets of query sequences were then compared to full GenBank using BLAST (Zhang et al., 2000). BLAST output was interpreted using a custom python function, which implements a lowest common ancestor (LCA) approach for taxonomic assignment similar to the strategy used by MEGAN (Huson et al., 2007). In brief, after the BLAST search we recorded the most significant matches to the reference database (yielding the top 10% bit-scores) for each of the query sequences. If only a single taxon was present in the top 10%, the query was assigned directly to this taxon. If more than one reference taxon was present in the top 10%, the query was assigned to the lowest taxonomic level that was shared by all taxa in the list of most significant hits for this query. Sequences for which the best BLAST hit had less than 97% identity to any sequence on GenBank, were considered non-target sequences and discarded. Sequences were then filtered for the presence of *Dreissena rostriformis bugensis* using GenBank sequences: EF080862, JX945980, EF080861, U47651, AF495877, EU484436, KJ881409. We quantified the level of contamination by examining the species detected in the single species positive samples. Low level contamination was found from Human and bacteria by applying a 0.2% threshold to all samples, this minor contamination was removed.

3.3.5 Data analysis

In order to normalise the sequencing depth of each sample from the metabarcoding data, the number of sequences assigned to *D. rostriformis bugensis* was divided by the total number of sequences from that sample. To facilitate comparison of the abundance across different samples, the normalised abundances (% of the total sequences) was normalized to 1. We used Generalized Linear Models (GLMs) to investigate 1) the influence of density, biomass and time on the DNA copy number since the start of the mesocosm experiment (before the mussels were removed from the tanks only), and 2) the influence of mussel density and distance from source population on the DNA copy number (qPCR) and read count (metabarcoding). Initial model tests showed all data sets were overdispersed, using quasi-Poisson and zero-inflated models did not resolve the dispersion and therefore a negative binomial GLM with a log link function was used. To check for model suitability, we checked the residual deviance of each model fitted a chi square distribution (see Supplementary Information I for GLM results). Finally, we assessed the relationship between DNA copy number and distance along the river using all seven sites included in the qPCR experiment, using a Spearman's Rank Correlation since the data were not normally

distributed. All calculations and visualizations of the data were done with the statistical program R (R-Core-Team 2017), with GLMs performed using the MASS package (Venables et al., 2002).

3.4 Results

3.4.1 qPCR assay development and validation in mesocosms

The LoD determined from the serial dilutions of tissue DNA was 1 x 10^{-4} ng/µl per reaction with both standard and qPCR, which equates to approximately 5 copies per reaction (See Supplementary Information I Fig. S3.1 for qPCR amplification gel images and Fig. 3.3 for amplification plot). We previously demonstrated that *D. r. bugensis* can be detected in mesocosms by PCR after four hours and at all three densities (Blackman et al., 2017c *in review*). This result was confirmed here using qPCR (Fig. 3.2). Rapid DNA accumulation was seen in the first two days of the mesocosm experiments, but this is followed by a depletion in DNA concentration until removal of mussels (Fig. 3.2). This pattern of DNA accumulation and depletion is highly consistent across all three density treatments (Fig. 3.2). Time since the beginning of the experiment (until removal of the mussels) had a significant effect on the DNA copy number in each mesocosm experiment when all density treatments were analysed together (GLM, $Z = -7.224$, $P < 0.0001$, AIC = 912.95). There is still some overdispersion with this model, however the overdispersion parameter indicates this is not substantial (overdispersion statistic = 1.3, Residual Deviance 91.142 on 70 df, χ^2 $P = 0.0457$). There is substantial overlap between the DNA copy number estimates for the three density treatments (Fig. 3.2) and neither density nor total biomass significantly explained the number of DNA copies detected in the qPCR (Density - $Z = -1.019$ $P = 0.308$, Total biomass - $Z = -0.776$ $P = 0.438$). DNA was still detected 24 hours after removal of the mussels, but not at the next sampling event at 7 days after removal (day 28) (Fig. 3.2). All blanks and mesocosm control tank samples were clear of target DNA throughout the experiment.

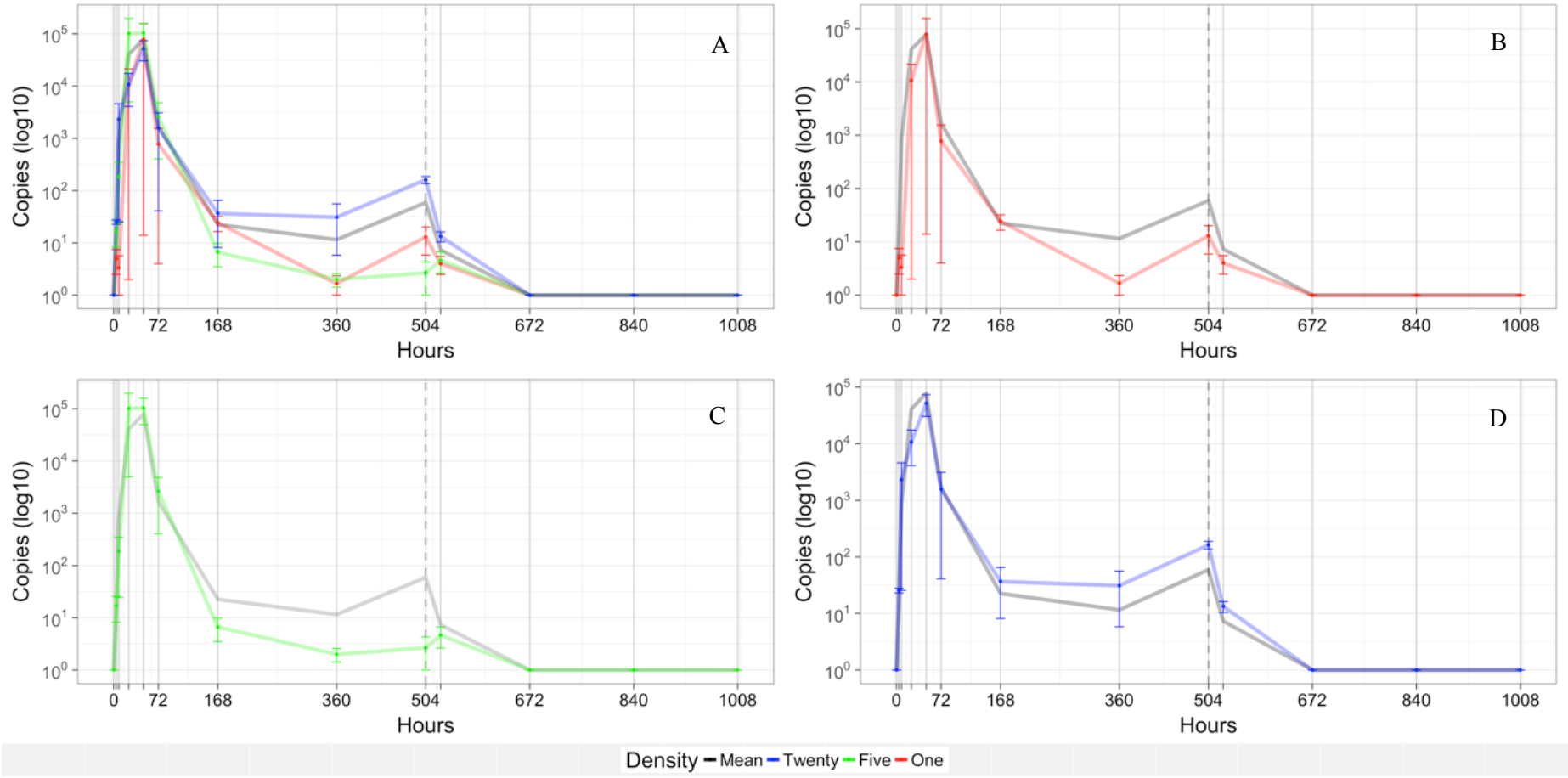


Figure 3.2: *Dreissena rostriformis bugensis* qPCR assay validation – mesocosm. Each graph shows the mean and standard deviation of DNA copy number recorded for each density treatment during the mesocosm experiment. The black line indicates the mean copy number of all three densities. A – all three densities (1, 5 and 20 specimens), B – 1 specimen, C – 5 specimens and D – 20 specimens.

3.4.2 Detection of *D. r. bugensis* by traditional and molecular approaches in the field

Kick sampling: The number of *D. r. bugensis* specimens varied from 14 at Wraysbury Weir, the site closest to the reservoir outfall, to a single specimen at Wraysbury Gardens, the site furthest from the outfall (Table 3.1). Detection rate was also highest at Wraysbury Weir (mussels detected in 4/6 sampling events). Mussels were detected in 1/6 samples at each of the other two sampling sites (Table 3.1)

Table 3.1 A - C: Comparison of methods for the detection of *D. r. bugensis* at three sites on the Wraysbury River. + indicates a positive detection and – indicates no detection for each technical replicate (i.e. three per sample for the eDNA methods and one for the kick sample). The numbers in parentheses in the Kick Sampling column correspond to the number of individuals found in each sample.

A. Wraysbury Weir (WW)

Sample No.	Kick sampling	PCR	qPCR	Metabarcoding
1	-	+ + +	+ + +	+ + +
2	+ (4)	+ + +	+ + +	+ + +
3	+ (4)	+ + +	+ + +	+ + +
4	-	+ + +	+ + +	+ + +
5	+ (3)	+ + +	+ + +	+ + +
6	+ (3)	+ + +	+ + +	+ + +

B. Wraysbury Bridge (WB)

Sample No.	Kick sampling	PCR	qPCR	Metabarcoding
1	-	+ + +	+ + +	+ + +
2	-	+ + +	+ + +	+ + +
3	-	+ + +	+ + +	+ + +
4	-	+ + +	+ + +	+ + +
5	+ (4)	+ + +	+ + +	+ + +
6	-	+ + +	+ + +	+ + +

C. Wraysbury Gardens (WG)

Sample No.	Kick sampling	PCR	qPCR	Metabarcoding
1	-	+ + +	+ + +	- + -
2	-	+ + +	+ + +	- - +
3	+ (1)	+ + +	+ + +	- - +
4	-	+ + +	+ + +	- - -
5	-	+ + +	+ + +	- - -
6	-	+ + +	+ + +	- - -

Targeted approaches: All six replicates from each of the field-collected samples were positive for quagga mussel with both PCR and qPCR (Table 3.1 and Fig. S3.1). Strong PCR products were produced for every replicate (n= 54) with PCR (Fig. S3.2). For both PCR and qPCR all negative and non-target tissue samples were clear of amplification for *D. r. bugensis* DNA. Figure 3.5A shows the DNA copy number (log10) for *D. r. bugensis* found at each of the sampling sites. Both distance from source and number of mussels found at the three sites, significantly influenced the DNA copy number, the strongest predictor however was distance (Density Z = -3.367, P = 0.000761, Distance Z = -17.223 P < 2e-16, AIC: 1392.1, Residual Deviance 57.621 on 51 df, χ^2 P = 0.244). A Spearman's Rank correlation between the DNA copy number at each of the seven sites screened by qPCR and distance from the source population showed a weak correlation (ρ = -0.25, P = 0.5948). However, this is likely affected by the low copy numbers found immediately downstream of the reservoir outfall (Fig. 3.3a). If this site is treated as an outlier and removed from the analysis, the correlation is significant (ρ = -1, P = 0.002778).

Passive approach: The total assigned read count passing quality control, after chimeric sequence removal was 4,928,265, of which 0.6% of reads (i.e. 29,570) were identified as *D. r. bugensis* after application of a 0.2% threshold. The run identified a further 281 OTUs in total including a range of macroinvertebrate and algae/diatom species (Supplementary information I, Fig. S3.4). The proportion of reads assigned to *D. r. bugensis* from each sample within each site varied greatly as seen in Figure 3.3B (n=54). The target species was detected in all samples from both Wraysbury Weir and Wraysbury Bridge sites, while detection at Wraysbury Gardens was much lower with only 3 of the 18 replicates having *D. r. bugensis* reads (Table 3.1). Furthermore, the raw number of *D. r. bugensis* reads was substantially lower with average read counts of 843 and 801 at Wraysbury Bridge and Wraysbury Weir, respectively and only 1 read on average at Wraysbury Gardens. The read count shows a strong relationship with both the number of mussels collected at the three sites and the distance from the source populations, with distance the strongest best predictor (Distance: z = -4.707, P < 0.0001, Density - z = -3.773, P = 0.0001, AIC = 258.64, Residual Deviance 49.696 on 51 df, χ^2 P = 0.820), in agreement with the qPCR results.

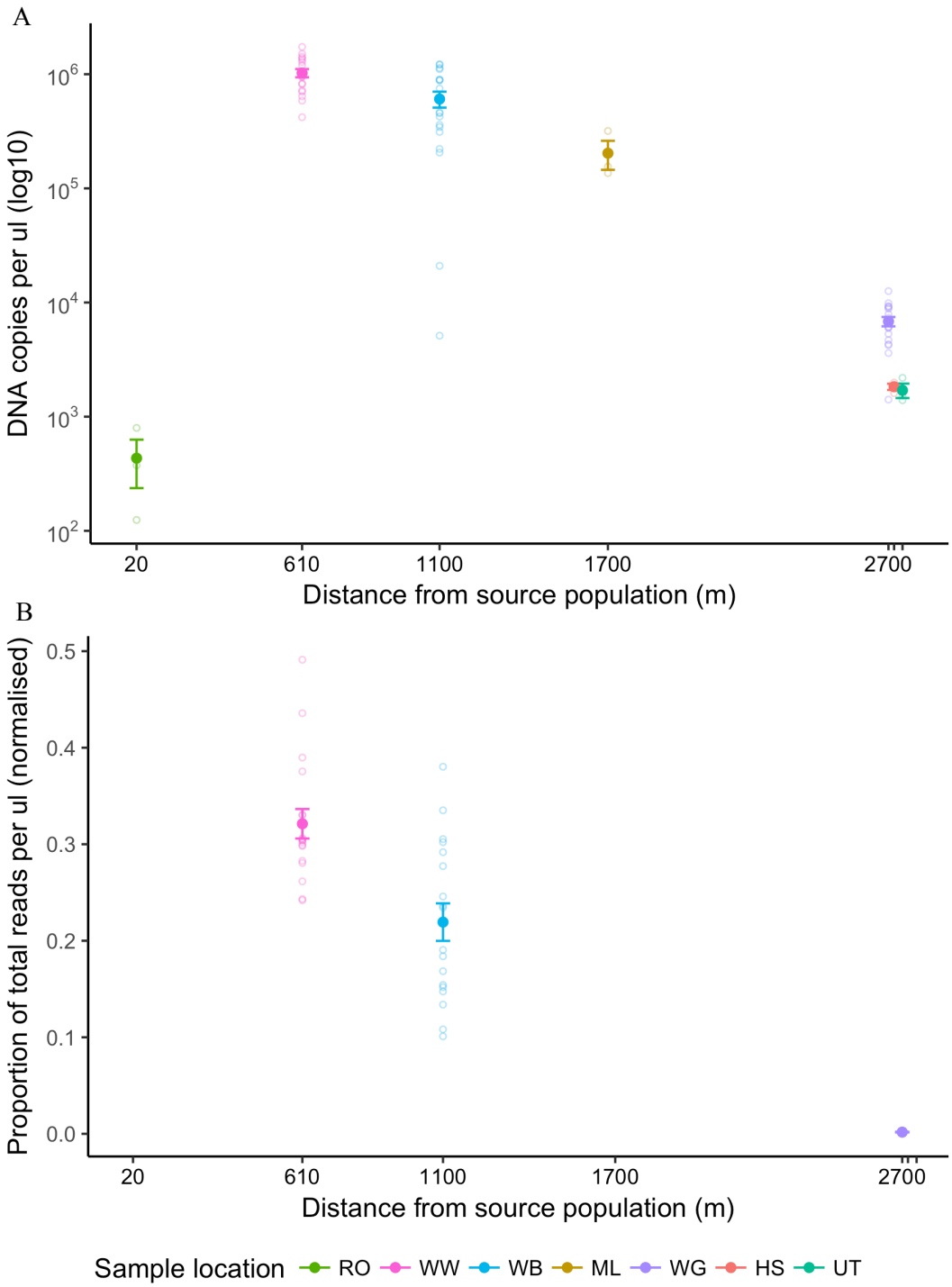


Figure 3.3: *Dreissena rostriformis bugensis* DNA signal recorded at field sites. A - qPCR DNA copy number per µl (log10). B - Metabarcoding read count per sample. The graph shows the read number for each site normalised over the total number of reads obtained from each sample using the metaBEAT pipeline.

3.5 Discussion

In this study, we developed and validated a dye-based qPCR assay for *D. r. bugensis* then compared targeted (PCR and qPCR) and passive (metabarcoding) approaches to traditional kick-net sampling for qualitative and quantitative detection in the field. In line with our first hypothesis, probability of detection was higher for all three eDNA methods than traditional methods. Detection probability was 100% in the field for both targeted approaches whereas metabarcoding was less sensitive than the targeted methods at the lowest quagga mussel density (detection probability 86%). Finally, both density and distance from the source population were significant predictors of read count (metabarcoding) and copy number (qPCR) indicating that both approaches can provide some indication of quagga mussel abundance in the field.

3.5.1 qPCR validation

Mesocosm validation provides an opportunity to test the limits of detection of eDNA assays and to investigate the relationship between density, time and DNA copy number (Dejean et al., 2011; Thomsen et al., 2012b; Blackman et al., 2017c *in review*). DNA was detected with qPCR at all three densities (1, 5 and 20 individuals) within 4 hours of being present, in agreement with previous standard PCR results (Blackman et al., 2017c *in review*). The limits of detection from serial dilutions of tissue DNA were also identical for standard and qPCR (1×10^{-4} ng/µl). More surprisingly, the mesocosm experiments highlighted unexpected dynamics of DNA in closed systems. Firstly, time since the start of the experiment was a significant predictor of DNA copy number, but rather than increasing over time or reaching a plateau, DNA accumulated quickly in the 48 hours then rapidly depleted, and this trend was replicated in all three density treatments. This could perhaps be explained by mussel behaviour. One of the adaptations which make *D. r. bugensis* such a successful invader is their ability to produce protein based byssal threads and thereby attaching themselves to substrate easily (Ricciardi et al., 1998; Karatayev et al., 2002; Aldridge et al., 2004; Timar and Phaneuf, 2009; Peyer et al., 2009), during the first 48 hours of the experiment the mussels are likely to be securing themselves to the new substrate, as all mussels were found to have done this when they were removed at the end of the experiment. The mussels are also filter-feeders, and were actively feeding throughout the experiment producing sloughed cells, faeces and pseudofaeces (Sansom and Sassoubre., 2017), this feeding behaviour is likely to be a major contributor to successful detection in the field, however when in a closed system, the

mussels behaviour is likely to also consume some of the eDNA they are producing, hence we see fluctuations in copy number over the course of the experiment. We hypothesize that after 48 hours, the degradation of DNA combined with the suspected consumption of eDNA via filter feeding exceeded production, leading to rapid depletion. The second surprising result from the mesocosm experiments was that we found no effect of density or biomass on DNA copy number. This contradicts several previous studies that have demonstrated a relationship between biomass or density and DNA copy number (Thomsen et al., 2012b; Takahara., et al 2012; Doi et al., 2015a; Doi et al., 2015b). Again, this could in part be due to the physiology and feeding behaviour of mussels, since eDNA is being consumed as well as produced. The complex pattern of eDNA accumulation and depletion together with the lack of discrimination between different density treatments, presents challenges for estimation of quagga mussel abundance. However, eDNA is unlikely to accumulate in the same way in natural (particularly lotic) environments, and this case highlights that mesocosm experiments may poorly reflect the true relationship between DNA copy number and mussel density in the field. It would be worthwhile to test whether this result is confirmed with other filter feeding species. More in keeping with previous studies, rapid degradation of eDNA was observed in mesocosm experiments, with a complete loss of signal by one week after mussel removal (Thomsen et al., 2012b).

3.5.2 Comparison of eDNA and kick-sampling for quagga mussel detection

The probability of detection for kick-net sampling was 33.3%, compared to 100% for targeted eDNA approaches and 86% for metabarcoding. The lower sensitivity for kick-net sampling is further highlighted by the failure to collect a *D. r. bugensis* specimen in any of the first three-minute kick-samples at each site. As demonstrated in a previous study, species in low abundance are likely to be missed with only 62% of Families detected in a 3 minute kick-sample (Furse et al., 1981), and therefore highlights the need for multiple replicates for detection of low abundance species if using kick-sampling. It should be noted that focussed established surveillance methods for *D. r. bugensis* prioritise the preferred habitat of mussels i.e. hard substrates such as boulders, cobbles and man-made structures (Aldridge et al., 2014), which is more likely to detect quagga mussels than the randomised method used here. However randomised kick sampling along a range of substrate types is the most widely used sampling method for freshwater macroinvertebrate monitoring, and therefore the most likely to detect new INNS.

Of the three molecular approaches compared in this study, we show that both targeted approaches were more sensitive for detecting *D. r. bugensis* than metabarcoding. Few studies have compared targeted approaches with metabarcoding, however our findings are similar to those of Simmons et al., (2015) and Harper et al., (2017), who showed that metabarcoding was less sensitive than targeted approaches for detecting bighead carp, *H. nobilis* and great crested newts, *Triturus cristatus*, respectively. Although in this study there was no increase in the number of detection of *D. r. bugensis* without applying a contamination threshold (0.2%), it would be pertinent to examine metabarcoding runs without applying thresholds at first, in order to identify any potential INNS with low read counts, with detection of priority INNS verified with a targeted assay. A significant advantage of metabarcoding is the added information it provides on non-target species. A further 281 OTUs were identified in this dataset, and although no additional INNS were detected, previous studies have demonstrated the power of this approach for detecting new and unexpected INNS (Blackman et al., 2017a). Metabarcoding also provides the opportunity to monitor changes in community composition as the result of INNS or other ecological stressors (Simmons et al., 2015; Hänfling et al., 2016), which is extremely promising for better understanding INNS impacts for more effective prioritisation and management. It could also be possible to improve the power in metabarcoding by increasing spatial sampling and future studies need to investigate how many samples are needed in order to maximise the probability of detection of rare species with metabarcoding

As mentioned previously, in the present study PCR and qPCR had identical performance for detection. This contradicts several studies that have indicated greater sensitivity of qPCR over PCR (Thomsen et al., 2012b; Nathan et al., 2014). It is possible that the sensitivity of the current qPCR assay could be increased even further by using a probe-based approach. However, given the already high sensitivity of the methods described, we believe they should be suitable for most monitoring needs. The approach chosen will depend on the specific requirements of the monitoring programme. If quantitative information is not required, standard PCR offers a quick and cost-effective approach for the detection of new INNS, requiring limited lab equipment and up to half the time for sample processing than qPCR (Davison et al., 2016). If quantitative information is required, for example when estimating population sizes and monitoring areas where eradication attempts have been carried out, qPCR will obviously be more appropriate, but it is worth noting that other studies have suggested that qPCR may be more prone to errors,

including false positives and false negatives, compared to standard PCR (Nathan et al., 2014; De Ventura et al., 2017).

3.5.3 Estimating quagga mussel abundance with eDNA

Results from field trials for both qPCR and metabarcoding demonstrate a significant decrease in DNA copy number or read count with increasing distance from the main source population. Kick-net sample data also indicates that mussel density decreases along the course of the river, with only one mussel found at Wraysbury Gardens, 2.7 km downstream of the reservoir. Quantitative data from qPCR and metabarcoding is particularly informative in this case as the concentration gradient suggests that the source population may be inferred in cases where it is unknown. The only site that did not fit this trend was the first site, 20m downstream of the outfall from Wraysbury reservoir, which had very low DNA copy number. This site is often littered with dead and empty *D. r. bugensis* shells, which could be due to a to decrease in the population of *D. r. bugensis* within Wraysbury Reservoir itself, and therefore concentrations of mussels are higher at sites further downstream. However, this is purely speculative.

Several studies have tried to quantify the effects of flow on the availability and transport of eDNA in lotic waterbodies (Deiner and Altermatt, 2014; Jane et al., 2015; Wilcox et al., 2016; Shogren et al., 2017). In particular one study highlighting rivers as a huge source for biodiversity assessment spanning both aquatic and terrestrial species (Deiner et al., 2016). However, there are still many answered questions regarding the factors which affect eDNA persistence and distribution in lotic systems. DNA degradation rates in aquatic ecosystems depend on a wide range of biotic and abiotic factors which are not yet fully understood (Barnes and Turner., 2014) and the varying rates in which eDNA disperses in riverine systems adds additional complexity. Broadly, there are two main processes which are known to influence the transport of DNA: firstly, the effect of stream discharge on transport (Jane et al., 2015), and secondly the fundamental structure of the river (Wilcox et al., 2016; Shogren et al., 2017). During this study, sampling took place at a period of increased flow (following heavy rain). In contrast to the findings of Jane et al., (2015), the DNA signal was not uniform along the stretch as expected by higher flows and increased mixing. It is therefore likely other factors were influencing the DNA signal across the river length. Shogren et al., (2017) in particular, summarised the elements most influential on the movement of DNA as “Transport, Retention and Resuspension”. As our data indicates distance

has a significant correlation with DNA copy number along the river stretch and it is highly likely that our data is documenting both a decrease in DNA being shed by a falling population size and also the effect of DNA being retained by the substrate in slower flow areas (slacks, glides and pools) and resuspended in higher velocity flows (riffles and runs). To what extent these factors are influencing the DNA signal is currently unclear, and further studies during different flow rates are needed in order to explore these variables further.

3.6 Conclusion

As the use of eDNA moves from research into application, the falling cost, growing number of species specific primers, development of laboratory and field protocols, the potential this method has for routine monitoring, and the detection of new and existing INNS is staggering. In this study, we have demonstrated a robust method for detection of *D. r. bugensis* via three molecular methods. The most effective approach in this study was the use of targeted primers, in agreement with similar studies. However, the development of species specific primers entails a lengthy development and validation process, which should not be under estimated and this approach misses unexpected INNS. Although having lower detection probability in the current and previous studies, metabarcoding has the obvious benefit of simultaneous multi-species detection. There are several possible ways in which the sensitivity of metabarcoding could be increased, for example using primers that are conserved in a narrower range of taxa, or increasing the spatial resolution of sampling. The optimal approach at present would be to carry out routine passive monitoring and verify any low read count detections with targeted assays. Certainly, both targeted and passive molecular approaches offer a greater ability to detect new and existing INNS than current methods, and would benefit current monitoring practices.

3.7 Supplementary Information I

Table S3.1: Sample sites used in this study: Site information for the points where eDNA samples were collected for each experiment on the River Wraysbury. * denotes samples used in experiment 2 where kick-net samples were not collected.

Site Name	Latitude	Longitude	Grid ref
Wraysbury reservoir outfall*	51.457536	-0.51827824	TQ0304374214
Wraysbury Weir	51.452369	-0.52052826	TQ0289873636
Wraysbury Bridge	51.448498	-0.52381396	TQ0267973201
Moor Lane*	51.443378	-0.52220432	TQ0280272634
Wraysbury Gardens	51.452369	-0.52052826	TQ0334071906
Hale Street*	51.435291	-0.51485463	TQ033371745
Upstream Thames confluence*	51.434893	-0.51465174	TQ0334671701

DNA Extraction method – Protocol for DNA extraction from filter papers modified from Brolaski et al., (2008)

Lysis solution 1 - 0.12µM Guanidine thiocyanate and 0.181 µM Tridocium phosphate

Lysis solution 2 – 5 µM sodium chloride, 0.5 µM Tris base, 4% SDS

Precipitation solution – 5 µM ammonium acetate, 0.12 alluminium ammonium sulphate dodecahydrate

Binding solution – 5 µM Guanidine HCl, 0.03 µM Tris HCl, 9% Isoproponol

Wash solution – 0.01 µM Tris HCl, 0.5 µM Sodium Chloride, 75% Ethanol

Elution Buffer – TE buffer

1. 1g 30mesh garnet beads, 1g fine sand into 7ml tube
2. Add filter paper
3. 925 µl Lysis solution 1 and 75 µl Lysis solution 2
4. Qiagen Tissue lyser 5 minutes, 30 bps
5. Centrifuge 4000g, 1min
6. Pipette off supernatant into clean 2ml tube
7. Add 250 µl Precipitation solution, vortex
8. Chill on ice for 5 mins
9. Centrifuge 10000g, 1min
10. Pipette off supernatant into clean 2ml tube
11. Add x1.5 volume of Binding solution, vortex
12. Pipette 650 µl into spin column, centrifuge 10000g, 1 min, discard flow through
13. Repeat step 12 until all solution has gone through spin column
14. Add 500 µl l of Wash solution, centrifuge 10000g, 1 min, discard flow through
15. Centrifuge spin column 10000g, 2 min
16. Place spin column in fresh collection tube
17. Add 100 µl of Elution buffer (TE, ddH_2O) leave for 5 minutes
18. Centrifuge 10000g, 1 min.

Generlized Linear mixed models

Table S3.2: *D. r. bugensis* field qPCR DNA copies GLM, mesocosm tests.

Model	AIC	Estimate Std. Error z value Pr(>|z|)	Residual deviance	Residual deviance P-val (1-pchisq)*
Hours + TotalBio	914.94	0.002236 -7.265 3.73e-13 *** 0.044271 0.078 0.938	91.141 on 69	0.03840271
Hours + Density	914.87	0.002233 -7.055 1.72e-12 *** 0.042593 -0.354 0.723	91.138 on 69	0.03842042
TotalBio	925.27	0.04715 -0.776 0.438	92.236 on 70	0.03873005
Hours	912.95	0.002235 -7.224 5.07e-13 ***	91.142 on 70	0.04567788
Density	924.91	0.04530 -1.019 0.308	92.205 on 70	0.03891364

Table S3.3: *D. r. bugensis* field qPCR DNA copies GLM, field trials comparrison

Model	AIC	Estimate Std. Error z value Pr(>|z|)	Residual deviance	Residual deviance P-val (1-pchisq)*
Distance+ Density	1392.1	0.000173 -17.223 < 2e-16 *** 0.027777 -3.367 0.000761 ***	57.621 on 51	0.2436637
Distance	1400.7	0.0001063 -23.78 <2e-16 ***	58.276 on 52	0.2555702
Density	1491.1	0.03279 5.102 3.37e-07 ***	67.462 on 52	0.07331911

Table S3.4: *D. r. bugensis* field metabarcoding reads GLM.

Model	AIC	Estimate Std. Error z value Pr(>|z|)	Residual deviance	Residual deviance P-val (1-pchisq)*
Distance+ Density	258.64	0.0009511 -4.707 < 2.51e-06 *** 0.0480304 -3.773 0.0001 ***	41.696 on 51	0.820419
Distance	301.86	0.0001565 -13.47 <2e-16 ***	54.989 on 52	0.3621245
Density	694.99	0.005852 18.99 <2e-16 ***	505.52 on 52	0 - overdispersed

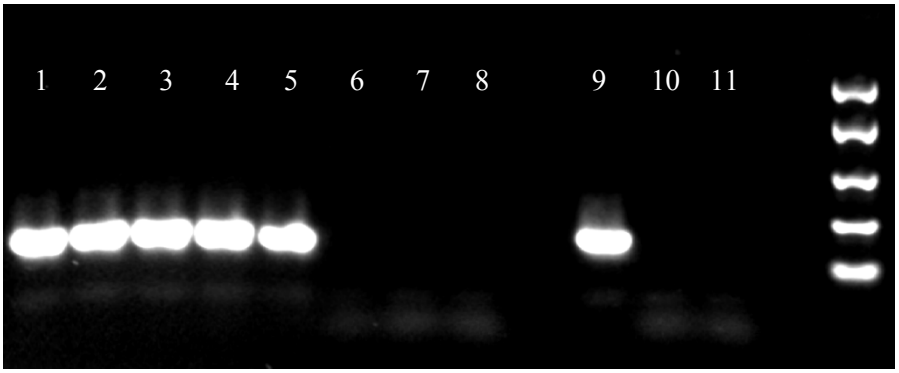


Figure S3.1: qPCR product visualised on an agarose gel to show amplification. Lane 1 – 1ng/µl, Lane 2 1:10 dilution, 0.1 ng/µl per reaction; Lane 3 1:100 dilution, 1 x 10^{-2} ng/µl per reaction; Lane 4 1:1000, ~1 x 10^{-3} ng/µl per reaction, Lane 5 1:10000, ~1 x 10^{-4} ng/µl per reaction, Lane 6 1:100000, ~01 x 10^{-5} ng/µl per reaction, Lane 7 1:1000000, ~1 x 10^{-6} ng/µl per reaction, Lane 8 1:10000000, ~01 x 10^{-7} ng/µl per reaction, Lane 9 is a positive tissue sample and 10 and 11 are PCR negative (ddH_2O) (Concentrations were measured on QuBit 2.0 prior to qPCR reaction and therefore Lanes 4-8 given are given as approximations only) The final lane is DNA EasyLadder I (Bioline, UK) with corresponding fragment size, the first and second band being 100bp and 250bp, respectively.

Figure S3.2. Agarose gel images from *D. r. bugensis* field samples collected from River Wraysbury. A – Wraysbury Weir, B – Wraysbury Bridge and C – Wraysbury Gardens. Blue highlights the successful detection of the target INNS above the required threshold (2/3) The final lane is DNA EasyLadder I (Bioline, UK) with corresponding fragment size.

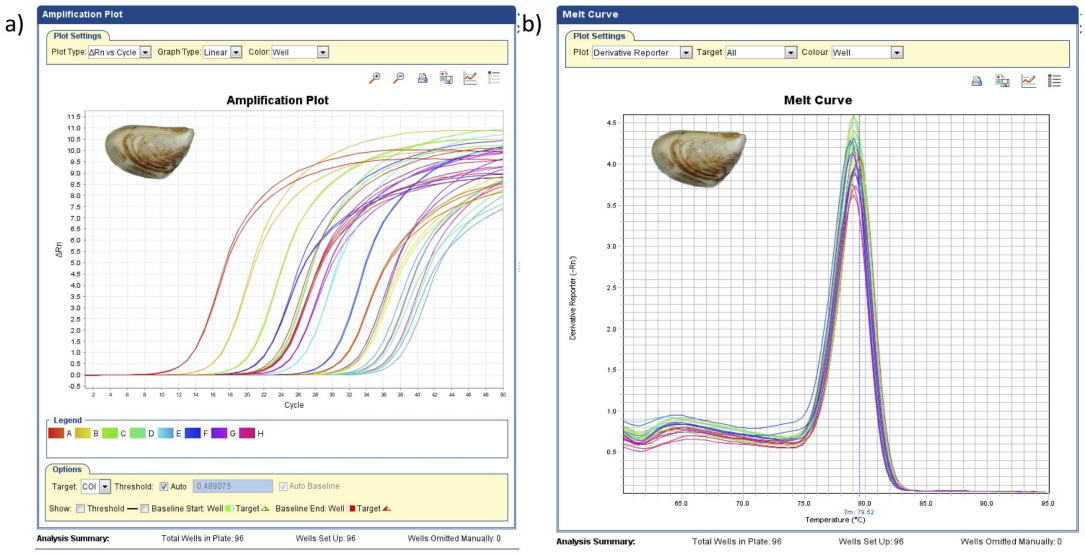


Figure S3.3: qPCR (SYBR green) amplification plot (a) and melt curve (b) for quagga mussel, *Dreissena rostriformis bugensis*. The amplification plot represents the accumulation of PCR product over the duration of the qPCR experiment. The Y axis is the change in relative fluorescence. Each coloured line on the plot represents a sample. The earlier in the qPCR experiment the curve increases from the baseline fluorescence, the greater the number of target DNA copies in the sample. The melt curve charts the change in fluorescence observed when double stranded DNA separates into single strands. Plotting the curve is a way of checking for reaction specificity. A single, clear peak (as shown here) demonstrates a highly specific reaction

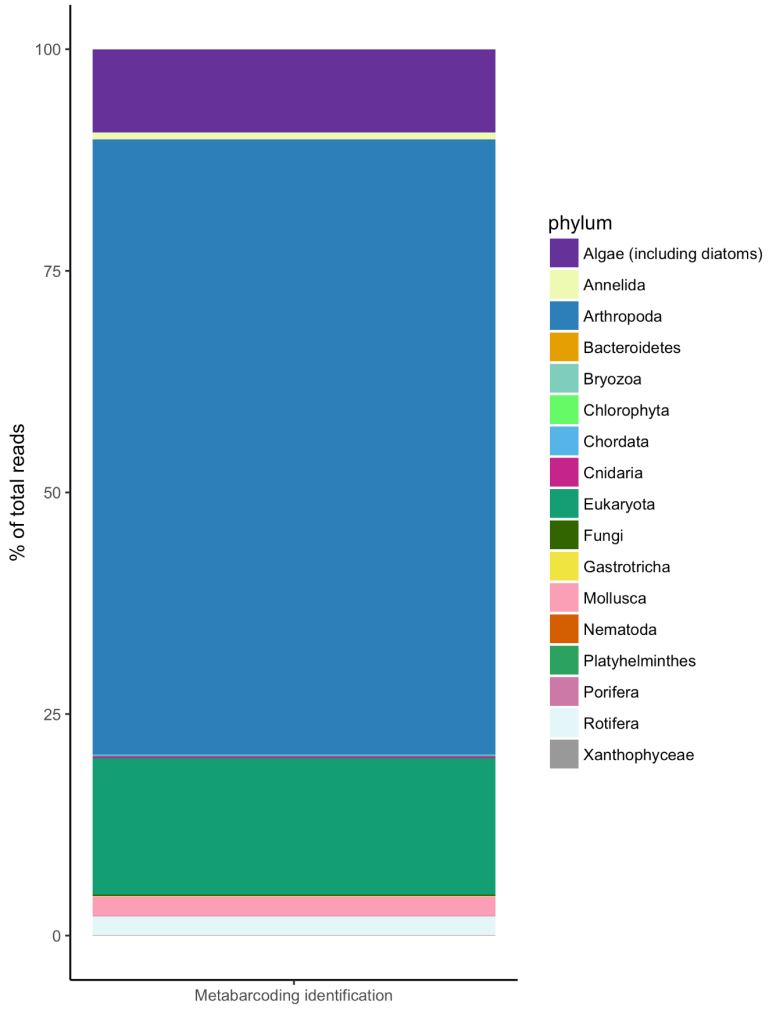


Figure S3.4: Other taxa detected from the metabarcoding analysis. A further 281 taxa were detected using the Leray et al (2013) primer set, excluding the target species *Dreissena rostriformis bugensis*. This stacked bar chart shows the taxa to Phylum/group level to indicate the range of species detected using the metabarcoding approach

3.8 Supplementary Information II – MIQE Checklist

Table S3.5: MIQE guidelines checklist for qPCR assay development and reporting

ITEM TO CHECK	*IMPORTANCE*	*CHECKLIST*
EXPERIMENTAL DESIGN		
Definition of experimental and control groups	E	Mesocosm experiment: 3 densities of *D. r. bugensis* (1, 5, and 20) with a tank with no specimens (control) Field experiment: 7 sites sampled on the R. Wraysbury with a control sample upstream of *D. r. bugensis* spread
Number within each group	E	Mesocosm control: n=1, Experimental: n=9, Total: n = 10 Field experiment control: n = 3, Experimental: n = 66
Assay carried out by core lab or investigator's lab?	D	Investigator's Lab
Acknowledgement of authors' contributions	D	Yes
SAMPLE		
Description	E	eDNA water sample
Volume/mass of sample processed	D	25ul
Microdissection or macrodissection	E	N/A

Processing procedure	E	Mesocosm: 200ml water samples were collected from each mesocosm at 4hrs, 8hrs, 24hrs, 7 days, 15 days, 21 days, 22 days, 28 days, and 42 days. Samples were vacuum filtered through sterile 47 mm diameter 0.45 µm cellulose nitrate membrane filters with pads (Whatman, GE Healthcare, UK) immediately after collection, using Nalgene filtration units (Thermo Fisher Scientific) in combination with a vacuum pump (15~20 in. Hg, Pall Corporation). Filtered samples were returned to the correct mesocosm_tank Field: 3 x 500ml water samples were collected at each sampling point and processed within 24 hours in the same way as the mesocosm samples. All samples were stored in petri dishes at –20 °C until DNA extraction. DNA extractions were carried out using a protocol modified from Bolaski et al., (2008) (for the full extraction protocol, see Supplementary Information I). DNA was eluted with 100 µl of Buffer AE. The DNA solution was stored in a 0.8ml microtube at –20 °C until PCR analysis.
If frozen - how and how quickly?	E	N/A
If fixed - with what, how quickly?	E	N/A
Sample storage conditions and duration (especially for FFPE samples)	E	All concentrated samples were stored in 0.8ml microtubes at –20°C.
NUCLEIC ACID EXTRACTION		
Procedure and/or instrumentation	E	We used a modified Bolaski et al., (2008) protol as documented in Supplementary information I
Name of kit and details of any modifications	E	We used a modified Bolaski et al., (2008) protol as documented in Supplementary information I

Source of additional reagents used	D	University of Hull
Details of DNase or RNAse treatment	E	N/A
Contamination assessment (DNA or RNA)	E	N/A
Nucleic acid quantification	E	Quantification was performed using a Qubit 2.0 following the manufacturer's instructions.
Instrument and method	E	
Purity (A260/A280)	D	
Yield	D	
RNA integrity method/instrument	E	N/A
RIN/RQI or Cq of 3' and 5' transcripts	E	N/A
Electrophoresis traces	D	N/A
Inhibition testing (Cq dilutions, spike or other)	E	N/A
REVERSE TRANSCRIPTION		
Complete reaction conditions	E	N/A
Amount of RNA and reaction volume	E	N/A
Priming oligonucleotide (if using GSP) and concentration	E	N/A
Reverse transcriptase and concentration	E	N/A
Temperature and time	E	N/A
Manufacturer of reagents and catalogue numbers	D	N/A
Cqs with and without RT	D	N/A

Storage conditions of cDNA	D	N/A
qPCR TARGET INFORMATION		
If multiplex, efficiency and LOD of each assay.	E	N/A
Sequence accession number	E	DQ840132.1
Location of amplicon	D	amplicon location: 196 - 384
Amplicon length	E	Including primers - 188 bp
In silico specificity screen (BLAST, etc)	E	The in-silico specificity screen was performed using Primer-BLAST. The *D. r. bugensis* primer pair, DRB1, amplified 29 published *D. rostriformis, D. bugensis* and *D. rostriformis bugensis* sequences in silico with no mismatches.
Pseudogenes, retropseudogenes or other homologs?	D	Not Found
Sequence alignment	D	N/A
Secondary structure analysis of amplicon	D	Not Checked
Location of each primer by exon or intron (if applicable)	E	N/A
What splice variants are targeted?	E	N/A
qPCR OLIGONUCLEOTIDES		
Primer sequences	E	DRB1_F(5'-GGAAACTGGTTGGTCCCGAT-3') DRB1_R (5'-GGCCCTGAATGCCCCATAAT-3')

RTPrimerDB Identification Number	D	Not Submitted
Probe sequences	D	N/A
Location and identity of any modifications	E	
Manufacturer of oligonucleotides	D	IDT Ltd
Purification method	D	HPLC
qPCR PROTOCOL		
Complete reaction conditions	E	Reactions were set up manually in specialist eDNA laboratory.
Reaction volume and amount of cDNA/DNA	E	Reaction volume is 25 µl, and amount of DNA is 2 µL
Primer, (probe), Mg++ and dNTP concentrations	E	N/A
Polymerase identity and concentration	E	We used Power Up SYBR Master Mix following manufacturer's instructions
Buffer/kit identity and manufacturer	E	We used Power Up SYBR Master Mix following manufacturer's instructions
Exact chemical constitution of the buffer	D	N/A
Additives (SYBR Green I, DMSO, etc.)	E	12.5 ul
Manufacturer of plates/tubes and catalog number	D	Microamp, Optical 96 Well Reaction Plate (10411785) & Microamp Optical Adhesive Film (10299204) (Applied Biosystems)

Complete thermocycling parameters	E	2 min at 50°C, 10 min at 95°C, and 55 cycles of 15 s at 95°C and 60 s at 60°C.
Reaction setup (manual/robotic)	D	We performed following the manufacturer's instructions.
Manufacturer of qPCR instrument	E	StepOne-Plus™ Real-Time PCR system (Applied Biosystems, Foster City, CA, USA)
qPCR VALIDATION		
Evidence of optimisation (from gradients)	D	Not Performed
Specificity (gel, sequence, melt, or digest)	E	The specificity of the primer was tested by sequencing PCR product by Macrogen
For SYBR Green I, Cq of the NTC	E	None
Standard curves with slope and y-intercept	E	Slope: Range -2.97 − -3.69, y-intercept: Range - 32.85−42.96
PCR efficiency calculated from slope	E	86.6−116%
Confidence interval for PCR efficiency or standard error	D	Standard error for PCR effciency = 1.68
r2 of standard curve	E	0.9906−0.99969
Linear dynamic range	E	Linear dynamic range from 124 to 1733209 DNA copies per reaction
Cq variation at lower limit	E	The positive signals were detected from the two of three wells for five copy template.
Confidence intervals throughout range	D	Not Checked
Evidence for limit of detection	E	Because five copies of target DNA was detected in at least two wells in each set of triplicates, we defined the limit of detection as 5 copies.

If multiplex, efficiency and LOD of each assay.	E	N/A
DATA ANALYSIS		
qPCR analysis program (source, version)	E	StepOne Software ver 2.0
Cq method determination	E	We performed according to default setting of Software above.
Outlier identification and disposition	E	
Results of NTCs	E	Three wells of no-template negative control were included in all qPCR plates and showed no amplification.
Justification of number and choice of reference genes	E	N/A
Description of normalisation method	E	We used standard curve methods.
Number and concordance of biological replicates	D	N/A
Number and stage (RT or qPCR) of technical replicates	E	Triplicate for each qPCR.
Repeatability (intra-assay variation)	E	N/A
Reproducibility (inter-assay variation, %CV)	D	N/A
Power analysis	D	N/A
Statistical methods for result significance	E	We performed according to default setting of Software above.
Software (source, version)	E	StepOne Software ver 2.0
Cq or raw data submission using RDML	D	Not Submitted

Chapter 4 Detection of a new non-native freshwater species by DNA metabarcoding of environmental samples - first record of *Gammarus fossarum* in the UK

4.1 Abstract

We report the discovery of a non-native gammarid, *Gammarus fossarum* (Crustacea, Amphipoda, Koch, 1836) in UK rivers. *Gammarus fossarum* is a common freshwater gammarid in many parts of mainland Europe, but was previously considered absent from the UK. *Gammarus fossarum* was detected in a number of UK rivers following DNA metabarcoding of a mini-barcode region of the COI gene in macroinvertebrate kick samples, and environmental DNA (eDNA) from water and sediment samples. Subsequent morphological analysis and standard DNA barcoding showed that the species could be reliably identified and separated from *Gammarus pulex* (Linnaeus, 1758), the most dominant and widespread native freshwater gammarid in the UK. Our data demonstrate extensive geographical coverage of *G. fossarum* in the UK, spanning distant river catchments. At present there is no data to confirm the likely origin of *G. fossarum's* introduction. Subsequent re-examination of historic archive material shows the species to have been present in the UK since at least 1964. This study is among the first to demonstrate the potential of eDNA metabarcoding for detection of new non-native species.

Note: This chapter has been published

Blackman, Rosetta, Drew Constable, Christoph Hahn, Andrew Sheard, Jessica Durkota, Bernd Hänfling, and Lori Lawson Handley. 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 (2):177–89.

4.2 Introduction

Amphipods are successful invaders in freshwater ecosystems, with many invasive non-native species (INNS) having been observed to adversely impact indigenous species within Europe over the last century (Bij de Vaate et al., 2002; Grabowski et al., 2007). The introduction of non-native amphipods may not only lead to displacement of native congeners (e.g. Dick and Platvoet 2000; MacNeil and Platvoet 2005; Kinzler et al., 2009), but may also impact on ecosystem structure and functioning (MacNeil et al., 2010b; Piscart et al., 2011; Constable and Birkby 2016) and introduce novel pathogens to newly colonised areas (Bacela-Spychalska et al., 2012).

Once non-native species are widely established, efforts to reduce their impacts are often problematic, hence management strategies are strongly focused on preventing introductions or spread (e.g. the "check, clean, dry" campaign in the UK). Early detection is key to such strategies, either to improve the success of eradication programs or to prevent further establishment and dispersal (Roy et al., 2014; Dejean et al., 2012). For freshwater macroinvertebrates, INNS detection methods typically rely on sampling programmes and morphological identification. However, the standard UK monitoring method for macroinvertebrates, a three minute kick sample, will typically recover 62% of families and 50% of species at a site (Furse et al., 1981). This can present considerable challenges when dealing with rare or elusive species. Morphological identification can also prove difficult when identifying taxonomically similar or cryptic species, or juvenile life stages, and is highly dependent on the taxonomical expertise of the investigator. Emerging molecular detection methods may provide significant benefit for detecting non-native species in aquatic environments (Darling and Mahon 2011; Lawson Handley 2015).

One new and rapidly developing method is the use of environmental DNA (eDNA) (Taberlet et al., 2012a; Taberlet et al., 2012b; Rees et al., 2014; Lawson Handley 2015), which refers to cellular or extracellular DNA that can be extracted directly from environmental samples without prior separation of taxa (Taberlet et al., 2012a). Environmental DNA has been successfully used in numerous studies to detect specific taxa using a targeted approach based on standard or quantitative PCR (Dejean et al., 2012; Dougherty et al., 2016). In an alternative approach, called "metabarcoding", entire species assemblages are analysed by PCR with broadly conserved

primers, followed by Next Generation Sequencing (NGS: see Lawson Handley 2015; Hänfling et al., 2016; Port et al., 2016; Valentini et al., 2016 for further detail). Environmental DNA metabarcoding has been successfully used in a small number of studies, for example, to describe entire communities of vertebrates (e.g. Lawson Handley 2015; Hänfling et al., 2016; Port et al., 2016; Valentini et al., 2016) and invertebrates (Deiner et al., 2016) from marine, lake and river samples. Metabarcoding has excellent potential as an early warning tool for detection of non-native species from samples collected from invasion pathways or natural/semi-natural habitats (Mahon and Jerde. 2016; Lawson Handley 2015). For example, the technique was recently used as an early detection method for screening ship ballast, and detected non-indigenous zooplankton in Canadian ports (Brown et al., 2016). Environmental DNA metabarcoding has also identified non-native fish species present in samples from the live bait trade (white perch, Morone americana (Gmelin, 1789) Mahon et al., 2014) and in river samples (northern snakehead, *Channa argus* (Cantor, 1842) Simmons et al., 2015). However the number of applications of metabarcoding for detection of non-native species has so far been limited.

In this paper we describe the detection of *Gammarus fossarum* (Koch, 1836), a newly recognised freshwater amphipod to the UK, using macroinvertebrate community and eDNA metabarcoding. The species was found in several UK rivers following a preliminary non-targeted sampling programme for macroinvertebrate communities based on metabarcoding of a 313 bp mini-barcode region of the cytochrome c oxidase subunit I (COI) gene, and was subsequently confirmed using a combination of morphological analysis and standard full-length COI DNA barcoding (via Sanger sequencing). This study demonstrates the power of eDNA metabarcoding for detection of non-native species in natural habitats.

4.3 Methods

4.3.1 Metabarcoding surveys

Field surveys were carried out in March 2015 within 8 UK river catchments (Figure 4.1 - Maps A - H, excluding E). At each site (n=65) environmental variables including water depth, width, substrate type and surrounding habitat were recorded. Three sample types were collected at each site: a three minute macroinvertebrate kick sample (Murray-Bligh 1999) for identification by microscopy analysis and high molecular weight DNA extraction from pools of individuals; and

water and sediment samples were collected for eDNA extraction. Two litres of water was sampled from the surface by collecting 4 × 500 ml from points across the river width using a sterile bottle. Sediment samples were collected from points across the river width using a trowel, and the material was placed in a 42 fluid oz. sterile Whirl-pak® bag (Cole-Palmer, Hanwell, London). All sampling equipment was sterilized in 10% commercial bleach solution for 10 minutes then rinsed with 10% MicroSol detergent (Anachem, UK) and purified water between samples. Sample bottles filled with ddH_2O were taken into the field and later filtered as sample blanks.

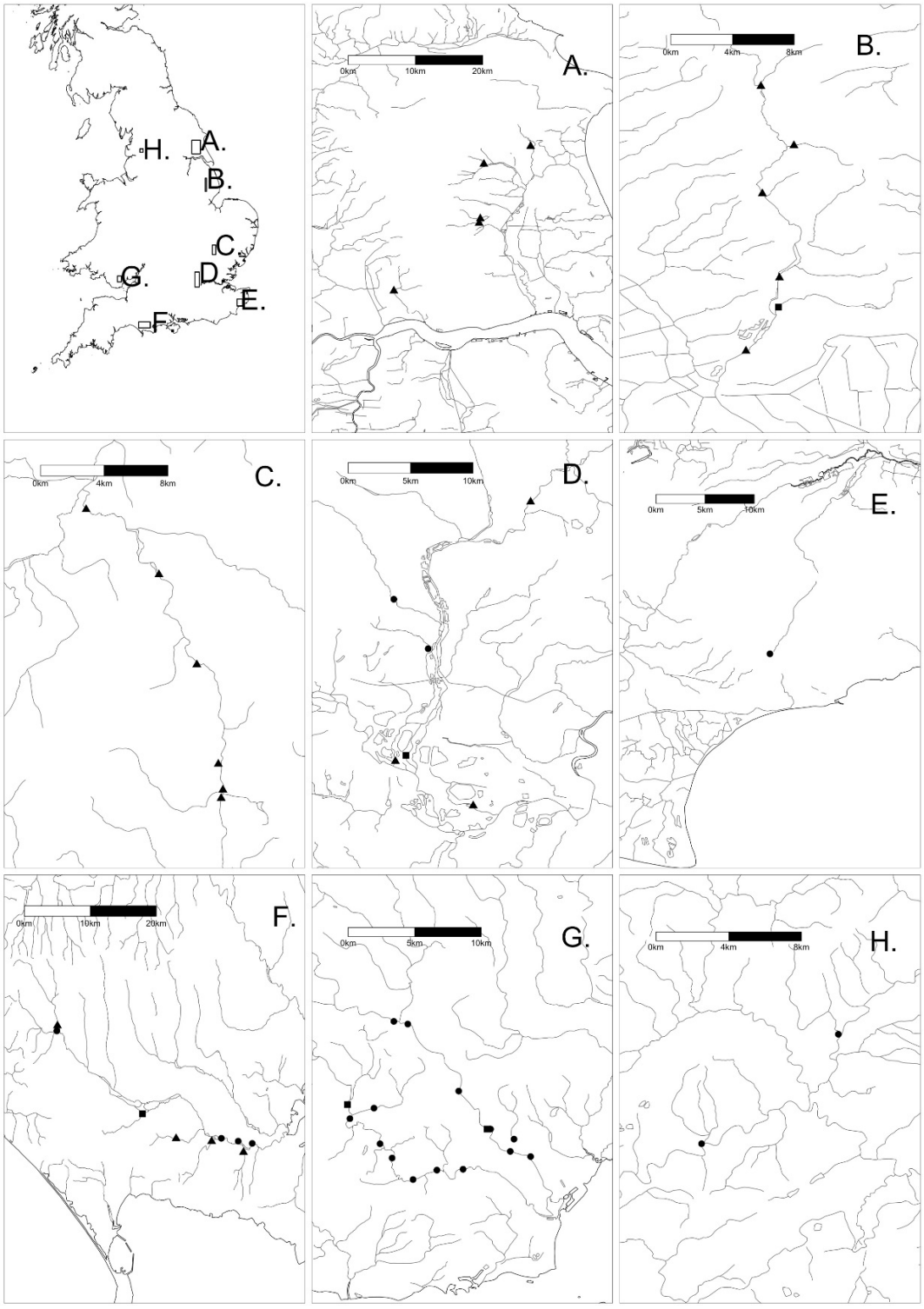


Figure 4.1. Distribution of Gammaridae species detected during this study. ● - *Gammarus fossarum*, ▲ - *Gammarus pulex* and ■ - both species present. A – River Hull, B - River Bain, C - River Cam, D - River Colne, E - Nailbourne, F - River Frome, G – Rivers Taff and Ely and H - River Ribble. See Supplementary Information I Table S4.1 for further site information.

4.3.2 Macroinvertebrate community sample processing

All macroinvertebrates from each kick sample were sorted and identified to the lowest taxonomic level possible, before being stored in sterile 50 ml falcon tubes filled with 100% ethanol. For DNA extraction, samples were dried to remove the ethanol and the entire macroinvertebrate community was lysed in a Qiagen Tissue Lyser® with Digisol (50mM Tris, 20M EDTA, 120mM NaCl and 1% SDS) (3 × 30 sec). Samples were then incubated overnight at 55°C with SDS and Proteinase K. DNA from a 200 µl subsample of the lysed tissue was extracted using the DNeasy Blood & Tissue Kit® (Qiagen, Hilden, Germany) according to the manufacturer's protocol.

4.3.3 Environmental DNA sample processing

Water samples were filtered within 24 hours through sterile 47 mm diameter 0.45 µm cellulose nitrate membrane filters and pads (Whatman, GE Healthcare, UK), using Nalgene filtration units attached to a vacuum pump. Sediment samples were stored at -20°C within 12 hours of sampling. The sample was defrosted, mixed and 200 ml of sediment placed in a sterile measuring cylinder with 500 ml of molecular grade water, then inverted 10 times and left to stand for 30s, the supernatant was then poured off into a sterile container. This procedure was repeated twice. Two hundred and fifty millilitres of the supernatant was then prefiltered through sterile 20 µm filter paper (Whatman, GE Healthcare, UK), and the filtrate subsequently filtered through 0.45 µm cellulose nitrate filters, as for the water samples. Filter papers were stored in sterile petri dishes at -20°C until extraction. Filtration blanks (2 L purified water) were run before the samples for each filtration run to test for contamination at the filtration stage (n=5). Filtration equipment was sterilized in 10% commercial bleach solution for 10 minutes then rinsed with 10% MicroSol detergent and purified water after each filtration.

Environmental DNA from both water and sediment samples was extracted using PowerWater® DNA Isolation Kit (MoBio Laboratories, Inc. Carlsbad, USA) following the manufacturer's instructions.

4.3.4 PCR, library preparation and sequencing

We chose to use COI for metabarcoding because this region has the broadest taxonomic coverage for macroinvertebrates in public sequence databases and is the most widely used DNA barcode for taxonomic discrimination in this group. A 313 bp fragment ("mini-barcode") was targeted using the primers described in Leray et al., (2013). For library preparation we used a nested tagging protocol, modified from the Illumina 16S two-step metabarcoding protocol (Illumina, 2011) as outlined in Kitson et al., (2015).

In the first step, PCRs were performed with modified versions of the primers jgHCO2198 TAIACYTCIGGRTGICCRAARAAYCA and mICOIintF GGWACWGGWTGAACWGTWT AYCCYCC (Leray et al., 2013). In addition to the standard primer sequence, primers included one of eight unique forward or 12 unique reverse 8-nucleotide Molecular Identification Tags (MID), plus a bridge site, which acts as a binding site for PCR 2 (see Kitson et al., 2015 for full details). PCRs were carried out in 25 µl volumes with MyFi High-Fidelity Taq (Bioline, UK) containing: 10 µM of each primer, and 2 µl of undiluted DNA template. PCRs were performed on an Applied Biosystems Veriti Thermal Cycler with the following profile: initial denaturation at 95°C for 1 min, followed by 45 cycles of denaturation at 98°C for 15s, annealing at 51°C for 15s and extension at 72°C for 30s, with a final extension time of 10 min at 72°C. This included PCR and filtering blanks (n=3 and n=5, respectively) and single species positives: *Triops cancriformis* (Bosc, 1801) (n=2) and *Harmonia axyridis* (Pallas, 1773) (n=2). PCR products were confirmed by gel electrophoresis on a 2% agarose gel stained with ethidium bromide. PCRs were carried out three times and then pooled. Pooled PCR products were then purified using the E.Z.N.A Cycle Pure Kit® (VWR International, Leicestershire).

In the second PCR step, Illumina adapters and additional forward and reverse MID tags were added in a second PCR with 10 µM of each tagging primer and 2 µl of purified PCR product. PCR settings were: initial denaturation at 95°C for 3 min, followed by 12 cycles of denaturation at 98°C for 20s, annealing at 72°C for 1 min and extension at 72°C for 5 mins, with a final extension time of 10 mins at 4°C (Kitson et al., 2015).

Samples were then classified into five categories based on the strength of band produced on ethidium bromide-stained agarose gels. Negative controls (including filtration blanks) produced no bands on the agarose gel so were categorised with samples with the lowest band strengths when being added to the library. All positive control (i.e. extracted tissue) samples were categorised as high band strength. Volumes of the samples were then pooled according 5 band strength categories: 10 µl for the lowest band strength, then decreasing volumes of 8 µl, 6 µl, 4 µl, and 2 µl for increasing band strength. The library was then pooled and cleaned using AMPure XP beads following the recommended manufacturer's protocol (Agencourt AMPure XP, Beckman Coulter Inc. US). The library was run at a 12 pM concentration on an Illumina MiSeq, at the in-house facility at the University of Hull, using the 2 x 300 bp V3 chemistry.

4.3.5 Specimen confirmation - microscopy and standard DNA barcode sequencing

Verification of the results from DNA metabarcoding was carried out using a combination of morphological identification and standard DNA barcoding (by Sanger sequencing).

Gammarus fossarum is a well-studied diverse species complex, which has three well established cryptic species (types A, B and C) with a further 36 - 53 different cryptic lineages being identified through phylogenetic studies (Weiss et al., 2014; Copilaş-Ciocianu and Petrusek, 2015). Species within this complex are known to differ in their ecology both in terms of their environmental requirements and geographic distributions (Copilaş-Ciocianu and Petrusek, 2015; Eisenring et al., 2016). The *G. fossarum* complex belongs to the G. pulex-group, which means it has small oval or kidney shaped eyes (less than twice as long as wide) and the pereopods 5-7 are armed with spines and few setae (Pinkster, 1972). Within the UK, these features alone would help to separate it from *G. duebeni, G. tigrinus* and *G. zaddachi*. It can be distinguished from all five known UK freshwater *Gammarus* residents by examining uropod III. In *G. fossarum* the ratio length of the endopod versus the exopod is about 0.5, whilst in the other five it is >0.5, typically 0.75 (see Figure 4.2B and 4.2D respectively). Another feature of *G. fossarum* is that only the inside margin of the exopod has plumose setae, whilst the other five have plumose setae on both inner and outer margins (see Figure 4.2C and 4.2E respectively). The latter feature should however be used with caution, as plumose setae on the outer margin of the exopod can show up in very old males of *G. fossarum* (Meijering, 1972).

A post hoc morphological examination of UK *Gammarus* specimens was carried out to confirm the presence of *G. fossarum*. Since the entire macroinvertebrate samples from the original sampling program had been lysed for metabarcoding, new specimens were collected by hand net from two catchments where *G. fossarum* was detected by metabarcoding in close proximity to previously sampled sites; River Taff, Wales (n=38) on 7/6/2016 and River Frome, England (n=39) on 27/6/2016. Additional, archived specimens obtained from the Nailbourne (Little Stour catchment), England (n=2) on 20/4/2013, were also analysed; (see Table 4.1 and Figure 4.1 Maps: E, F and G). Collected individuals were then subject to morphological examination and identified using Karaman and Pinkster (1977), Eggers and Martens (2001) and Piscart and Bollache (2012).

Microscopic identification was carried out on all specimens collected for morphological confirmation. Both *G. fossarum* (n=37) and *G. pulex* (n=1) were identified from individuals collected from the River Taff and only *G. fossarum* (n=39) was found in a sample from the River Frome. Standard DNA barcoding was performed on some of the individuals identified morphologically as *G. fossarum* (n=3) and *G. pulex* (n=1) from the River Taff, and *G. fossarum* from the Nailbourne (Little Stour catchment) (n=2). DNA was extracted using the DNeasy Blood & Tissue Kit® (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The full length COI DNA barcoding fragment was amplified (Folmer et al., 1994) using the following protocol: PCRs were performed in 25 µl volumes with MyTaq (Bioline, UK), 10 µM of each primer and 2 µl of DNA template. The PCR profile consisted of: initial denaturation at 95°C for 1 min, followed by 35 cycles of denaturation at 95°C for 15s, annealing at 50°C for 15s and extension at 72°C for 10s, with a final extension time of 10 min at 72°C. PCR products were checked on agarose gels and commercially sequenced using HCO2198 (Macrogen Europe, Amsterdam, Netherlands).

4.3.6 Bioinformatics

Processing of Illumina read data and taxonomic assignment were performed using a custom bioinformatics pipeline (metaBEAT, v.0.97.7-global; see Github reference 1) as described previously (Hänfling et al., 2016), with minor modifications. For each sample, raw Illumina sequences were filtered to retain only read pairs containing the expected forward/reverse in-line barcode combination (perfect matches only) using the program process_shortreads from the Stacks v1.20 program suite (Catchen et al., 2013) and subsequently quality trimmed using the

program Trimmomatic v0.32 (Bolger et al., 2014). Specifically, read quality was assessed across 5 bp sliding windows starting from the 3'-end, and reads were clipped until the per window average read quality reached a minimum of phred 30. Any reads shorter than 100 bp after the quality clipping were discarded. To remove PCR primers and spacer sequences the first 30bp of the reads was clipped off. Remaining sequence pairs were merged into single high quality reads using the program FLASH v1.2.11 (Magoč and Salzberg 2011). For any read pairs not merged successfully, only the forward read was retained for downstream analyses. Sequences were clustered at 97% identity using vsearch v1.1 (see Github reference 2). Any clusters represented by less than three sequences were excluded from further analyses, as these likely represent sequencing error. Each of the remaining distinct sequence clusters was collapsed to a single representative sequence (aka centroid). Only centroid sequences of the expected length as determined by the primers (313 bp +- 5%) were retained for downstream analyses. To obtain a final set of non-redundant (nr) queries for taxonomic assignment, centroid sequences across all samples were clustered globally at 97% identity using vsearch v1.1. The global set of nr queries was subjected to a BLAST (Zhang et al., 2000) search (blastn) against a custom reference database consisting of gammarid sequences from Weiss et al., (2014) and two CO1 sequences from *T. cancriformis* (GenBank accession numbers EF189678.1 and JX110644.1) and *H. axyridis* (accession numbers KU188381.1 and KU188380.1), respectively. Taxonomic assignment was performed using a lowest common ancestor (LCA) approach. In brief, after the BLAST search the algorithm identifies the most significant matches to the reference database (top 10% bit-scores) for each of the query sequences. If only a single taxon is present in this list of matches then the query is assigned directly to this taxon. If more than one taxon is present, the query is assigned to the lowest taxonomic level that is shared by all taxa in the list. Queries yielding best BLAST matches below a bit-score of 80 or with less than 85% identity were binned as 'unassigned'. To assure full reproducibility of our analyses we have deposited the entire workflow in an additional dedicated Github repository (see Github reference 3). To reduce the possibility of false positives based on our single species positive samples and in order to obtain a conservative estimate of the distribution of *G. fossarum* in the UK, we only report *G. fossarum* as present at a given site if it was supported by at least 1% of the total quality trimmed reads per sample.

4.3.7 Phylogeny

Phylogenetic analysis was performed to further confirm the identity of the putative *Gammarus* sp. sequences obtained as part of the current study. We downloaded a previously published CO1 dataset (Weiss et al., 2014; Copilaş-Ciocianu and Petrusek 2015) from Genbank, comprising 89 sequences of *G. fossarum*, six *G. pulex* (Linnaeus, 1758) sequences and a single sequence each from four further outgroup species (*G. balcanicus* (Schaferna, 1922), *G. glabratus* (Hou & Li, 2003), *G. roeselii* (Gervais, 1835) and *G. tigrinus* (Sexton, 1939) (Radulovici et al., 2009; Hou et al., 2011; Feckler et al., 2012; Weiss et al., 2014). This set of previously published sequences was extended by the sequences obtained via standard full-length DNA barcoding and mini-barcode metabarcoding. Prior to phylogenetic analysis we extracted the most abundant sequence, i.e. haplotype, from each sample from the initially obtained 97% sequence clusters assigned to *G. fossarum* and *G. pulex*, respectively. Nucleotide sequences of *G. fossarum* and *G. pulex* used in the phylogenetic analysis were deposited in Genbank (GenBank accession KY464959 - KY464977). Phylogenetic analysis was performed in the Reprophylo environment (Szitenberg et al., 2015). In brief, sequences were aligned using the program MAFFT v7.123b (Katoh and Standley 2013) and the alignment was trimmed using the program trimAl v1.2rev59 (Capella-Gutiérrez et al., 2009). Maximum-likelihood tree inference was performed using RAxML v8.0.12 (Stamatakis 2014). The full, detailed analysis is provided as Jupyter notebook in the dedicated Github repository (Github reference 3), which also contains the alignment underlying the phylogenetic tree and further supplementary information.

4.3.8 Comparison of data from eDNA/DNA and microscopy analysis

A correlation was performed to compare the Gammaridae abundance data generated from the kick sample microscopy analysis and the DNA/eDNA metabarcoding. Specifically, the relationship between DNA/eDNA data (read count) and data from microscopy analysis (biomass calculated from average Gammaridae specimen weight) was investigated by calculating Pearson's Correlation Coefficient in R v3.1.3 (R Core team 2015). Note that *G. fossarum* and *G. pulex* sequencing data have been combined here as the species were not distinguished during the initial morphological determination.

4.4 Results

4.4.1 Metabarcoding survey

The total sequence read count passing quality control, before removal of chimeric sequences, was 4,290,271. We quantified the level of possible contamination using sequence information from single species positive samples, which enabled us to choose a suitable threshold level (1% of total sample reads) for filtering and removal of low level contamination. This conservative threshold is comparable to recent, similar studies (e.g. Hanfling et al., 2016; Port et al., 2016). After applying this threshold, over the 195 samples the total read count was 933,457.

Gammarus fossarum was detected in 28 sites in total: 25 via metabarcoding, 1 site by morphological identification, 1 site by standard DNA barcoding and 1 site by morphological identification and DNA barcoding (See Table 4.1 and Supplementary Information I Table S4.1). Of the 25 metabarcoding samples, *G. fossarum* was found in: 25 DNA macroinvertebrate samples, 8 water eDNA samples and 9 sediment eDNA samples. *G. pulex* was detected in 27 of the sites in the metabarcoding DNA macroinvertebrate samples only and a single site using Sanger sequencing.

A full breakdown of gammarid sequences per sample and proportion of gammarid biomass per sample are included in Supplementary Information I Table S4.1. A further 36 freshwater macroinvertebrate families were detected by metabarcoding: data from these non-gammarid species form part of a wider macroinvertebrate data set which is being analysed separately and will be published elsewhere.

The average read count of the samples with gammarid species present was 3512. At those sites the proportion of *G. fossarum* reads per sample ranged from 1.68 - 100% in the macroinvertebrate DNA, 1.67 - 55.35% in the water eDNA and 1.59 - 18.05% in sediment eDNA samples (Supplementary Information I Table S4.1). Similarly, *G. pulex* reads ranged from 1.65 - 97.41% in the DNA macroinvertebrate samples. There was a significant positive correlation between the percentage of *Gammarus* biomass in the sample, and the percentage of *Gammarus* sequence reads

(Pearson's r = 0.747, df = 46, P = 1.098 x 10^{-9}, Supplementary Information I Figure S4.1). Importantly, *Gammarus* sequences were detected when gammarids constituted as little as 2.6% of the total biomass (Supplementary Information I Table S4.1).

4.4.2 Verification of Gammarus fossarum by microscopy

Gammarus fossarum was not identified morphologically in any samples surveyed in March 2015 prior to metabarcoding. Of the 38 gammarid specimens recovered from the River Taff on 7/6/2016, 37 *G. fossarum* morphological identifications were made. Adult males ranged between 8-12 mm (n=21) and adult females 7-10 mm (n=15). Four females were ovigerous. The other gammarid specimen encountered was a male *G. pulex* (13 mm). Of the 39 gammarid specimens collected from the River Frome on 27/6/2016, all were identified as *G. fossarum* morphologically. Adult males of this population ranged from 8-11.5 mm (n=24) and adult females 7-9 mm (n=15). Again, four ovigerous females were recorded. The relative abundance of size distribution in the two sampled populations can be seen in the Supplementary information I (Figure S4.2). The two individuals collected from the Nailbourne on 20/4/2013 were not verified using microscopy as the specimens were too heavily damaged for morphological identification.

The size ranges encountered for *G. fossarum* fall within the expected range for the species, with Goedmakers (1972), Pinkster (1972), Karaman and Pinkster (1977) and Piscart and Bollache (2012) reporting that the largest males typically reach 14-15 mm.

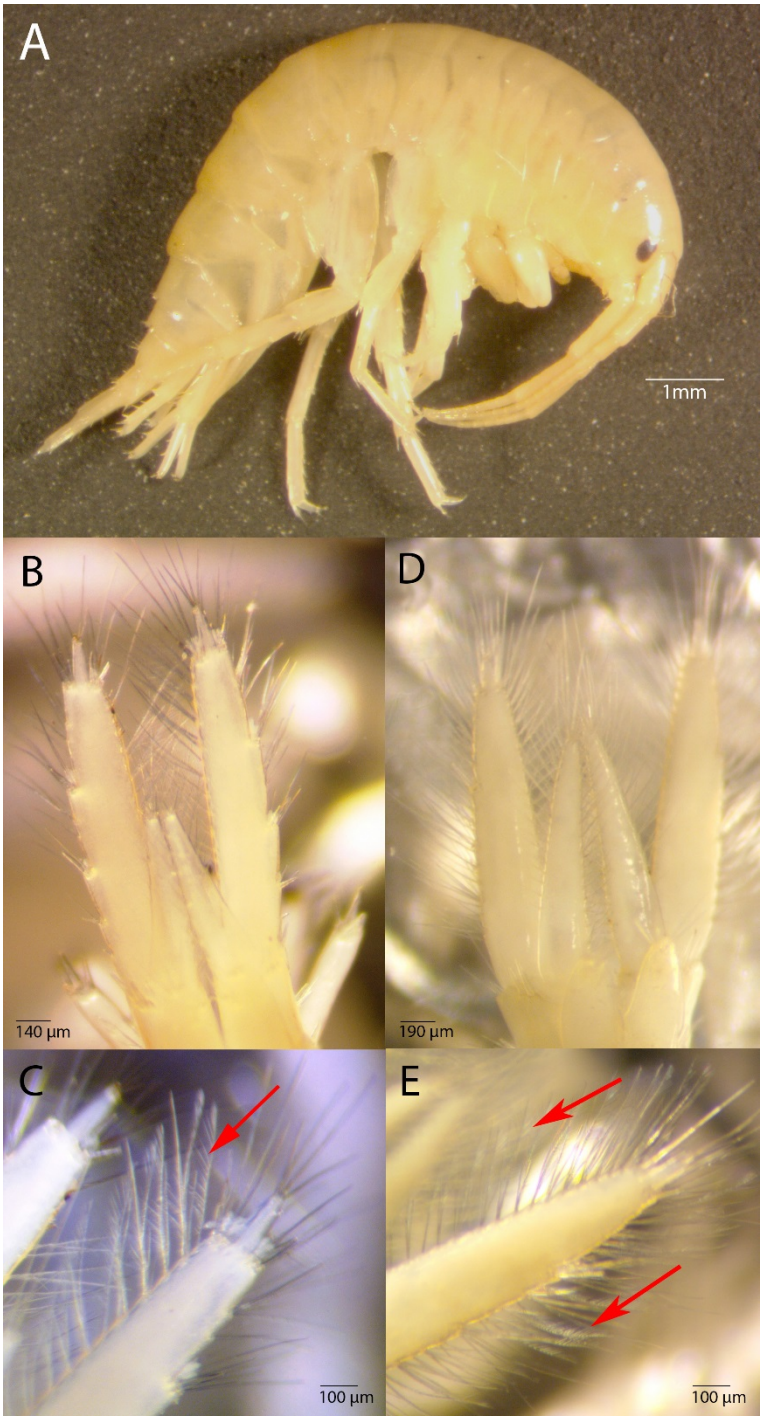


Figure 4.2. Picture of *Gammarus fossarum* found in the River Taff, UK, 7/6/2016, A) male adult specimen, B) male uropod III and C) male plumose hairs on inside of exopod of uropod III (→); and picture of male *Gammarus pulex* features for comparison D) uropod III and E) plumose hairs on inner and outer edge of exopod of uropod III (→) (Photographs by D. Constable).

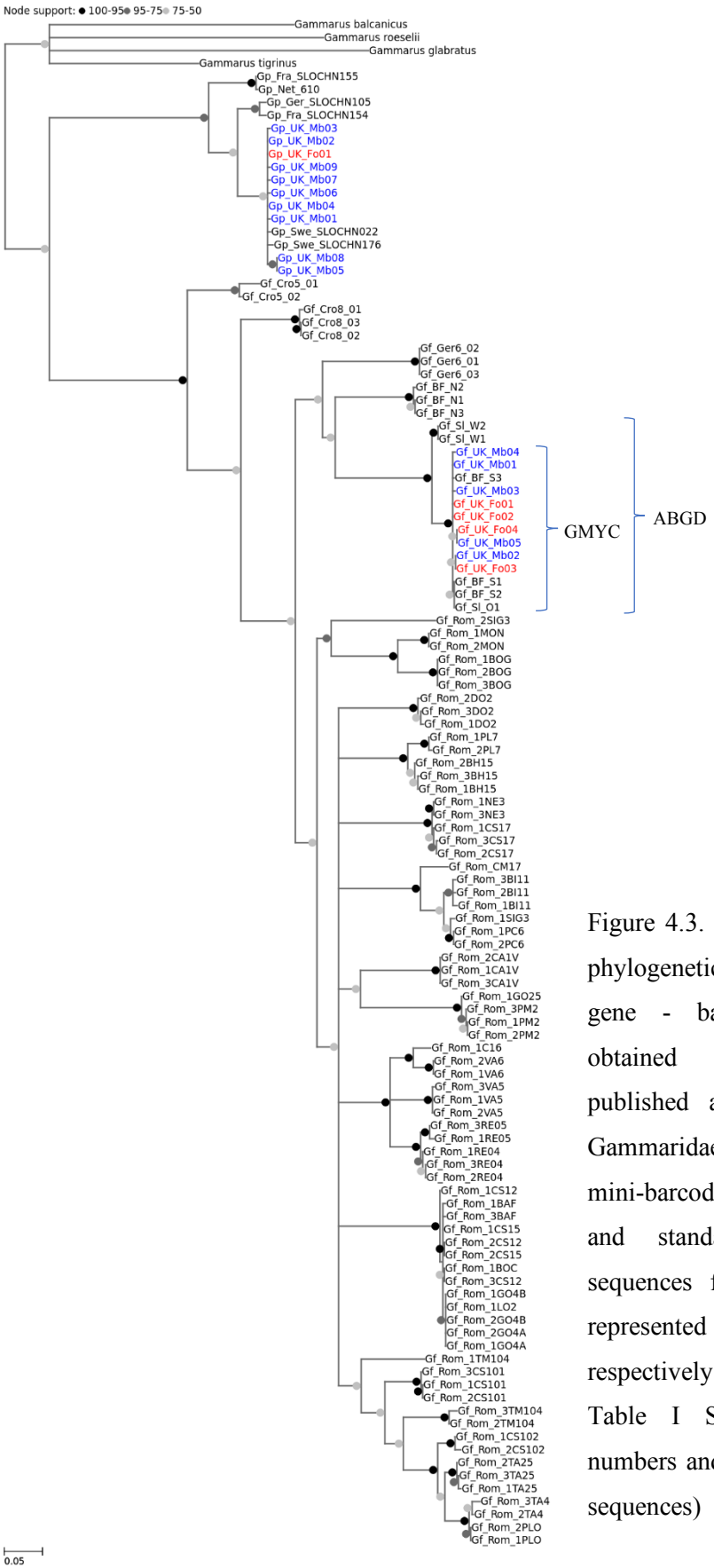


Figure 4.3. Maximum likelihood phylogenetic tree for the *COI* gene - based on sequences obtained from previously published and newly obtained Gammaridae sequences. The mini-barcode (metabarcoding) and standard *COI* barcode sequences from this study are represented in blue and red, respectively. (See supplementary Table I S4.2, for accession numbers and origin of individual sequences)

4.4.3 Verification of Gammarus fossarum by DNA barcoding

Morphological identifications were confirmed by DNA sequencing for specimens collected from the River Taff (n=4): 3 specimens of *G. fossarum* and a single *G. pulex*. The individuals collected from the Nailbourne (n=2) were also both identified as *G. fossarum* using subsequent DNA barcoding (see Table 4.1).

4.4.4 Phylogeny

The phylogeny (Figure 4.3) is congruent with the findings of the morphological identification. The *G. cf. fossarum* and *G. cf. pulex* sequences cluster with their respective lineages (identified in Weiss et al., 2014; Copilaş-Ciocianu and Petrusek 2015). *Gammarus fossarum* sequences obtained by both metabarcoding and standard DNA barcoding show little divergence and cluster together in the phylogeny, indicating closely related sequences. The *G. fossarum* sequences obtained in the current study group with high statistical support within Clade 11, as defined using the distance based Automatic Barcode Gap Discovery (ABGD) approach in Weiss et al., (2014). Sequences further group in a subclade with samples from south-western Germany, Southern Black Forest and Eastern Sauerland in Germany, i.e. clade 14, as delineated using the tree-based GMYC in Weiss et al., (2014). Aligning the UK *G. fossarum* specimens within Clade 11 confirms previous studies which show this clade to be the most widely distributed across Europe within the species complex (Copilaş-Ciocianu and Petrusek 2015; Weiss and Leese 2016).

Table 4.1. Specimen identification and identification method for morphologically identified and DNA barcoded specimens. (*Specimens collected from the River Frome were subject to morphological identification only. **Specimens collected from Nailbourne were DNA sequenced only due to damaged specimens).

					G. fossarum		*G. pulex*	
Unique ID	Catchment	Site Name	Latitude	Longitude	Microscopy	DNA sequencing	Microscopy	DNA sequencing
DC003	Taff	Forest Farm Country Park	51.516	-3.242	✓	✓		
DC004	Taff	Forest Farm Country Park	51.516	-3.242			✓	✓
DC005	Taff	Forest Farm Country Park	51.516	-3.242	✓	✓		
DC006	Taff	Forest Farm Country Park	51.516	-3.242	✓	✓		
DC007-045	Frome	East Stoke	50.681	-2.185	✓*			
JD001	Nailbourne	Adj Saint Ethelburga well	51.126	1.087		✓**		
JD002	Nailbourne	Adj Saint Ethelburga well	51.126	1.087		✓**		

4.5 Discussion

Non-targeted detection by direct and environmental DNA metabarcoding has the potential to revolutionise early warning systems for non-native species, but this utility of the new technology has so far been demonstrated only a limited number of times (Mahon et al., 2014; Brown et al., 2016). In this study, *G. fossarum*, a newly recognised non-native species for the UK, was detected during the course of a wider metabarcoding survey of macroinvertebrate communities. The identification of *G. fossarum* was subsequently confirmed by microscopy and standard DNA barcoding. The sequences generated from this study indicate that the UK populations of G. fossarum sampled here fall within the previously identified Clade 11, sensu Weiss et al., (2014), of this highly diverse species complex (Figure 4.3). Importantly this is the most widely distributed clade within the *G. fossarum* complex (Weiss et al., 2014; Copilaş-Ciocianu and Petrusek 2015; Weiss and Leese 2016).

Gammarus fossarum was found in seven distant river catchments within the UK, indicating a widespread distribution (Figure 4.1). Initial detection of *G. fossarum* was made using non-targeted metabarcoding of macroinvertebrate DNA, water eDNA and sediment eDNA samples. Of the sites where *G. fossarum* was detected using this method (n=25), *G. fossarum* was detected in all 25 DNA macroinvertebrate samples (100%), in 8 of water (32%) and 9 sediment (36%) samples. The lower detection of *G. fossarum* in eDNA samples compared to macroinvertebrate samples is not surprising due to the dilution of eDNA and effects of flow on DNA availability in lotic systems.

At 23 of the 28 sites (including post hoc samples) where *G. fossarum* was present it was the only Gammaridae species detected. This suggests it is not only widespread in the UK but could also be the dominant gammarid in some locations, possibly even having displaced the native *G. pulex* locally. With the new species discovery, recent re-examination of historical archived gammarid samples was undertaken from available Environment Agency and Natural History Museum (NHM), London, collections. Material from the Environment Agency had overlooked records of G. fossarum dating back to 2005 from the River Len, Maidstone, Kent (51.2619°N, 0.56451°E) whilst re-examination of material from the NHM revealed the earliest record to date, 1964 from

the River Darent, Kent. This shows that *G. fossarum* has remained undetected and overlooked by conventional means for a substantial length of time.

Gammarus fossarum is indigenous and widespread in mainland Europe, and typically inhabits springs and upper reaches of mountainous streams, with *G. pulex* being more dominant in lower river sections (Nijssen 1963; Goedmakers 1972; Karaman and Pinkster 1977; Chen et al., 2012). This distribution pattern is linked to *G. fossarum's* comparative preference for shallower streams and higher current velocities, and its reduced tolerance of low dissolved oxygen conditions (Meijering 1971; Peeters and Gardeniers 1998). It may also be found in middle sections of rivers and is able to coexist with *G. pulex* (Janetzky 1994; Piscart and Bollache 2012; Copilaş-Ciocianu et al., 2014). In such areas of coexistence, *G. fossarum* will often occupy faster flowing areas where vegetation is sparse or absent, and *G. pulex* will be found near marginal shore zones, with reduced currents and rich vegetation growth (Karaman and Pinkster 1977). The distributions of *G. fossarum* in this study covered a range of habitats, mainly lowland rivers (altitude <90 m) with the exception of the Nailbourne spring, adjacent to Saint Ethelburga Well and Maiden Newton on the Upper Frome, with altitudes of 106 m and 109 m, respectively (see Supplementary information I Table S4.1). The river depths at *G. fossarum* locations were shallow, seldom reaching more than 20 cm. It is important that further exploration of UK upland systems is undertaken as the sites surveyed for this study were mostly lowland, and at this stage are an indication of habitat suitability rather than preference for *G. fossarum* in the UK. Of our five study sites where *G. fossarum* and *G. pulex* co-existed, all had a mean depth >20 cm and featured both fast and slow currents as well as vegetative marginal areas, however there appears to be no other pattern in the distribution of sites where both species were found to co-exist. Four of the five sites were from the metabarcoding samples, the percentage read count for both species varied substantially, hence no species dominance can be inferred from this data (see Supplementary Information I Table S4.1).

Gammarus fossarum is the third non-native freshwater gammarid to be found in the UK within the last six years, following the discoveries of *Dikerogammarus villosus* in 2010 (MacNeil et al., 2010a) and *Dikerogammarus haemobaphes* in 2012 (Aldridge 2013). The record is rather unforeseen, and the species has not been included on the UK's non-native species watch list with more focus being placed on Ponto-Caspian species that have invaded western Europe (Gallardo

and Aldridge 2014). A detailed risk assessment of the threat that *G. fossarum* poses to native Gammaridae within the UK does not currently exist; further research into how *G. pulex* and *G. fossarum* co-exist within UK habitats should be carried out to decide if this action is warranted. However, the importance of this discovery as a new non-native species to the UK should not be overlooked as it has important implications for future ecological assessments.

4.6 Conclusion

In conclusion, we detected a newly recognised non-native species to UK fauna using non-targeted DNA metabarcoding, and confirmed its presence using microscopy and standard DNA barcoding. It is well known that the effectiveness of INNS control or management relies heavily upon early detection (Lodge et al., 2006; Vander Zanden et al., 2010). In future, for other species, non-targeted monitoring of high risk invasion pathways using eDNA may ensure that early eradication or containment are possible management options (Davis 2009; Hulme, 2009; Jerde et al., 2011; Thomsen et al., 2012b; Lawson Handley 2015). It is important that future research should now focus on establishing the true distribution, ecology and potential implications of *G. fossarum* within the UK, as well as exploring how the non-targeted eDNA metabarcoding approach can be used to detect non-native species.

4.7 Supplementary Information I

Table S4.1: Specimen identification, identification method and site information for metabarcoding samples, with presence of *G. fossarum* and *G. pulex*. *G. fossarum* percentage reads of total sample in brackets.

									G. fossarum			***G. pulex***
Unique ID	Site	Catchment	Width (m)	Depth (cm)	Altitude (m)	Latitude	Longitude	Sample *Gammarus* biomass (%)	Invert DNA	Water eDNA	Sediment eDNA	Invert DNA
RB005	Waddington	Ribble	4	20	51	53.87147	-2.41597	60.13	✓ 99.88		✓ 12.06	
RB010	Duddle Brook Confluence	Ribble	3.3	45	25	53.81131	-2.53073	39.95	✓ 98.85			
RB011	Great Chesterford Road Bridge	Cam	5.5	15	35	52.06213	0.19238	38.09				✓ 75.85
RB013	Whittlesford	Cam	8	40	22	52.11554	0.157535	52.2				✓ 19.6
RB015	B1383 Road Bridge	Cam	3	20	55	52.00317	0.212139	55.11				✓ 15.77
RB016	Road Bridge Debden Water	Cam	1.8	4	65	51.98293	0.214767	95.32				✓ 55.59
RB017	Water Lane, Newport	Cam	2	10	58	51.98786	0.216394	51.43				✓ 3.22
RB018	Hauxton Mill	Cam	7.5	25	10	52.1541	0.090515	17.26				✓ 7.84

RB019	Cottage Pasture	Hull	3	10	7	53.9033	-0.45552	46.06					✓ 18.14
RB021	Driffield Showground	Hull	12	40	14	53.99691	-0.44733	84.81					✓ 66.9
RB023	Kelk Beck - Railway Bridge	Hull	5	30	10	54.02722	-0.34051	82.97					✓ 97.41
RB025	Mill Beck - Town End Lane	Hull	3.5	25	12	53.77969	-0.65316	12.99					✓ 48.87
RB026	Scorborough	Hull	3	20	10	53.89559	-0.45827	65.68					✓ 39.26
RB027	Blackweir	Taff	35	27.33	10	51.49468	-3.196	53.88	✓ 93.5				
RB028	Llandaff Cathedral	Taff	27.7	32.67	20	51.49876	-3.21817	44.54	✓ 89.63				
RB029	Whitchurch	Taff	4	11.7	30	51.50841	-3.21413	98.2	✓ 100	✓ 5.45	✓ 1.67		
RB030	Radyr Weir	Taff	20.7	37.33	30	51.51642	-3.24203	76.33	✓ 99.52				
RB031	Taffs Well	Taff	22.7	35	30	51.54616	-3.27434	44.63	✓ 19.8				
RB033	Pontypridd	Taff	25	50	57.62	51.59878	-3.32984	11.11	✓ 1.68				

RB034	Taffs Well	Taff	20	31.7	65	51.60086	-3.34498	7.07	✓11.35				
RB035	St Fagans	Ely	8	23	10	51.48471	-3.2696	99.45	✓ 100	✓ 8.45	✓ 7.49		
RB036	St Georges	Ely	7.7	15	20	51.48404	-3.29809	99.71	✓ 99.91	✓ 4.35			
RB037	Peterston-Super-Ely	Ely	11	24.33	20	51.47665	-3.32409	94.83	✓ 99.66	✓ 13.53	✓ 1.59		
RB038	Bryn Farm	Ely	8.3	26.67	30	51.49356	-3.34691	99.23	✓ 100	✓ 2.9	✓ 2.17		
RB039	Pont Tal-Y-Bont	Ely	5.5	24	30	51.50474	-3.36006	98.65	✓ 100	✓ 4.15	✓ 5.72		
RB040	Pont Y Clun	Ely	9.7	16.67	40	51.5245	-3.3928	88.2	✓ 98.76	✓ 1.67	✓ 1.63		
RB041	Cross Inn	Ely	4	13	50	51.53272	-3.36666	99.64	✓ 99.87	✓ 55.35			
RB042	L'Oreal	Ely	3.8	18	50	51.53552	-3.39558	17.49	✓ 76.32			✓ 5.09	
RB043	Biscathorpe Ford	Bain	3	30	73	53.34668	-0.15229	58.3				✓ 53.11	
RB044	Hemingby	Bain	5.3	26	40	53.25127	-0.15089	30.83				✓ 15.99	

RB046	Dalderby	Bain	6	20	21	53.17648	-0.13417	8.65				✓ 4.02
RB048	Goulceby	Bain	2	15	61	53.29403	-0.12005	52.2				✓ 11.61
RB049	Haltham	Bain	2.3	30	14	53.15043	-0.1353	6.14	✓ 5.06			✓ 5.38
RB050	Wharfe Lane	Bain	15	20	7	53.11136	-0.16688	8.43				✓ 1.65
RB051	Moigne Combe	Frome	4.5	19.47	30	50.68402	-2.31985	61.66				✓ 90.55
RB052	Seven Stars	Frome	2.3	22	16	50.68158	-2.24321	86.09				✓ 67.81
RB053	Wool Bridge	Frome	18	29.58	20	50.68424	-2.22218	78.28	✓ 95.46			
RB054	Luckford Bridge	Frome	3.4	30.4	8	50.67222	-2.1745	38.45				✓ 61.69
RB055	Holme Bridge	Frome	14.3	41.1	10	50.67985	-2.15564	39.41	✓ 99.67			
RB056	West Stafford Rectory	Frome	4	20	51.48	50.70538	-2.39222	91.87	✓ 94.58			✓ 4.14
RB057	Maiden Newton - Hooke	Frome	6	39.22	90	50.77655	-2.57719	86.92	✓ 99.63		✓ 18.05	

RB058	Maiden Newton - Frome	Frome	4.5	30	109.53	50.78105	-2.57583	12.1				✓ 1.36
RB059	Colne Bk	Colne	4.2	24.33	30	51.54139	-0.49943	84.72	✓ 99.29			
RB060	Bushey Mill Lane	Colne	6.7	48.33	55	51.6705	-0.38036	72.05				✓ 73.66
RB061	Colne Brook, Horton Brook	Colne	4	30	19.19	51.44245	-0.53752	17.58				✓ 14.65
RB062	Gerards Cross	Colne	3.9	20.67	49	51.58491	-0.5393	91.46	✓ 97.41		✓ 2.15	
RB063	Staines Moor	Colne	5	40	15	51.44766	-0.52523	29.54	✓ 3.04			✓ 18.85
RB064	Ashford Common	Colne	4.5	40	12.4	51.40334	-0.44708	2.64				✓ 3.89

Table S4.2 Information on specimens from own and published studies that were used in the phylogenetic tree.

Sequence ID	Species	Genbank accession	Reference
Gammarus balcanicus	*Gammarus balcanicus*	JF965914.1	Hou et al., (2011)
Gammarus glabratus	*Gammarus glabratus*	EF570307.1	Hou et al., (2009)
Gammarus roeselii	*Gammarus roeselii*	JF965952.1	Hou et al., (2011)
Gammarus tigrinus	*Gammarus tigrinus*	FJ581688.1	Radulovici et al., (2009)
Gf_BF_N1	*G. fossarum*	KF521829.1	Weiss et al., (2013)
Gf_BF_N2	*G. fossarum*	KF521828.1	Weiss et al., (2013)
Gf_BF_N3	*G. fossarum*	KF521827.1	Weiss et al., (2013)
Gf_BF_S1	*G. fossarum*	KF521832.1	Weiss et al., (2013)
Gf_BF_S2	*G. fossarum*	KF521831.1	Weiss et al., (2013)
Gf_BF_S3	*G. fossarum*	KF521830.1	Weiss et al., (2013)
Gf_Cro5_01	*G. fossarum*	KF521806.1	Weiss et al., (2013)
Gf_Cro5_02	*G. fossarum*	KF521805.1	Weiss et al., (2013)
Gf_Cro8_01	*G. fossarum*	KF521804.1	Weiss et al., (2013)
Gf_Cro8_02	*G. fossarum*	KF521803.1	Weiss et al., (2013)
Gf_Cro8_03	*G. fossarum*	KF521802.1	Weiss et al., (2013)
Gf_Ger6_01	*G. fossarum*	KF521822.1	Weiss et al., (2013)
Gf_Ger6_02	*G. fossarum*	KF521821.1	Weiss et al., (2013)
Gf_Ger6_03	*G. fossarum*	KF521820.1	Weiss et al., (2013)
Gf_Rom_1BAF	*G. fossarum*	KR061776.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1BH15	*G. fossarum*	KR061778.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1BI11	*G. fossarum*	KR061781.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1BOC	*G. fossarum*	KR061784.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1BOG	*G. fossarum*	KR061785.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1C16	*G. fossarum*	KR061788.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1CA1V	*G. fossarum*	KR061789.1	Copilas-Ciocianu & Petrusek, (2015)

Gf_Rom_1CS101	*G. fossarum*	KR061792.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1CS102	*G. fossarum*	KR061795.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1CS12	*G. fossarum*	KR061797.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1CS15	*G. fossarum*	KR061800.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1CS17	*G. fossarum*	KR061802.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1DO2	*G. fossarum*	KR061805.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1GO25	*G. fossarum*	KR061842.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1GO4A	*G. fossarum*	KR061808.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1GO4B	*G. fossarum*	KR061810.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1LO2	*G. fossarum*	KR061812.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1MON	*G. fossarum*	KR061813.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1NE3	*G. fossarum*	KR061815.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1PC6	*G. fossarum*	KR061817.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1PL7	*G. fossarum*	KR061838.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1PLO	*G. fossarum*	KR061840.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1PM2	*G. fossarum*	KR061843.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1RE04	*G. fossarum*	KR061820.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1RE05	*G. fossarum*	KR061823.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1SIG3	*G. fossarum*	KR061825.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1TA25	*G. fossarum*	KR061827.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1TM104	*G. fossarum*	KR061830.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_1VA5	*G. fossarum*	KR061833.1	Copilas-Ciocianu & Petrusek, (2015)

Gf_Rom_1VA6	*G. fossarum*	KR061836.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2BH15	*G. fossarum*	KR061779.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2BI11	*G. fossarum*	KR061782.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2BOG	*G. fossarum*	KR061786.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2CA1V	*G. fossarum*	KR061790.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2CS101	*G. fossarum*	KR061793.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2CS102	*G. fossarum*	KR061796.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2CS12	*G. fossarum*	KR061798.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2CS15	*G. fossarum*	KR061801.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2CS17	*G. fossarum*	KR061803.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2DO2	*G. fossarum*	KR061806.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2GO4A	*G. fossarum*	KR061809.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2GO4B	*G. fossarum*	KR061811.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2MON	*G. fossarum*	KR061814.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2PC6	*G. fossarum*	KR061818.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2PL7	*G. fossarum*	KR061839.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2PLO	*G. fossarum*	KR061841.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2PM2	*G. fossarum*	KR061844.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2RE04	*G. fossarum*	KR061821.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2SIG3	*G. fossarum*	KR061826.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2TA25	*G. fossarum*	KR061828.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2TA4	*G. fossarum*	KR061846.1	Copilas-Ciocianu & Petrusek, (2015)

Gf_Rom_2TM104	*G. fossarum*	KR061831.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2VA5	*G. fossarum*	KR061834.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_2VA6	*G. fossarum*	KR061837.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3BAF	*G. fossarum*	KR061777.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3BH15	*G. fossarum*	KR061780.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3BI11	*G. fossarum*	KR061783.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3BOG	*G. fossarum*	KR061787.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3CA1V	*G. fossarum*	KR061791.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3CS101	*G. fossarum*	KR061794.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3CS12	*G. fossarum*	KR061799.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3CS17	*G. fossarum*	KR061804.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3DO2	*G. fossarum*	KR061807.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3NE3	*G. fossarum*	KR061816.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3PM2	*G. fossarum*	KR061845.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3RE04	*G. fossarum*	KR061822.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3RE05	*G. fossarum*	KR061824.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3TA25	*G. fossarum*	KR061829.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3TA4	*G. fossarum*	KR061847.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3TM104	*G. fossarum*	KR061832.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_3VA5	*G. fossarum*	KR061835.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Rom_CM17	*G. fossarum*	KR061819.1	Copilas-Ciocianu & Petrusek, (2015)
Gf_Sl_O1	*G. fossarum*	KF521835.1	Weiss et al., (2013)
Gf_Sl_W1	*G. fossarum*	KF521833.1	Weiss et al., (2013)

Gf_Sl_W2	*G. fossarum*	KF521834.1	Weiss et al., (2013)
Gf_UK_Fo01	*G. fossarum*	KY464960	current study
Gf_UK_Fo02	*G. fossarum*	KY464961	current study
Gf_UK_Fo03	*G. fossarum*	KY464962	current study
Gf_UK_Fo04	*G. fossarum*	KY464963	current study
Gf_UK_Mb01	*G. fossarum*	KY464973	current study
Gf_UK_Mb02	*G. fossarum*	KY464974	current study
Gf_UK_Mb03	*G. fossarum*	KY464975	current study
Gf_UK_Mb04	*G. fossarum*	KY464976	current study
Gf_UK_Mb05	*G. fossarum*	KY464977	current study
Gp_Fra_SLOCHN154	*Gammarus pulex*	JF965941.1	Hou et al., (2011)
Gp_Fra_SLOCHN155	*Gammarus pulex*	JF965942.1	Hou et al., (2011)
Gp_Ger_SLOCHN105	*Gammarus pulex*	JF965940.1	Hou et al., (2011)
Gp_Net_610	*Gammarus pulex*	EF570334.1	Hou et al., (2007)
Gp_Swe_SLOCHN022	*Gammarus pulex*	JF965939.1	Hou et al., (2011)
Gp_Swe_SLOCHN176	*Gammarus pulex*	JF965943.1	Hou et al., (2011)
Gp_UK_Fo01	*Gammarus pulex*	pending	current study
Gp_UK_Mb01	*Gammarus pulex*	pending	current study
Gp_UK_Mb02	*Gammarus pulex*	pending	current study
Gp_UK_Mb03	*Gammarus pulex*	pending	current study
Gp_UK_Mb04	*Gammarus pulex*	pending	current study
Gp_UK_Mb05	*Gammarus pulex*	pending	current study
Gp_UK_Mb06	*Gammarus pulex*	pending	current study
Gp_UK_Mb07	*Gammarus pulex*	pending	current study
Gp_UK_Mb08	*Gammarus pulex*	pending	current study
Gp_UK_Mb09	*Gammarus pulex*	pending	current study

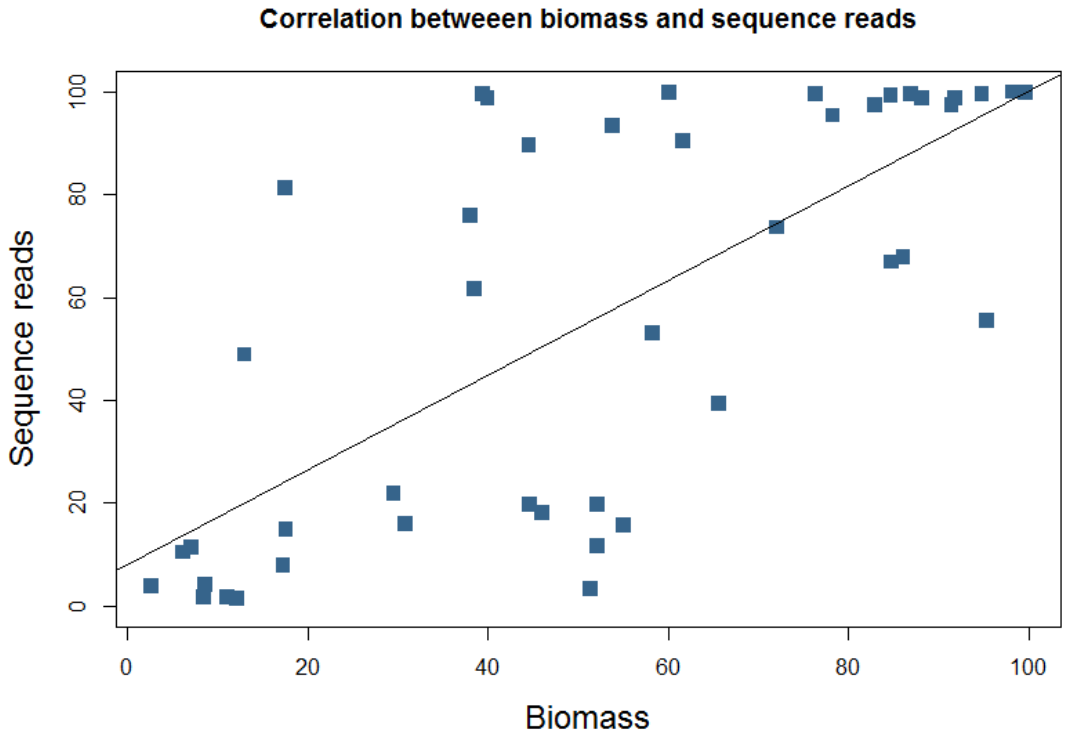
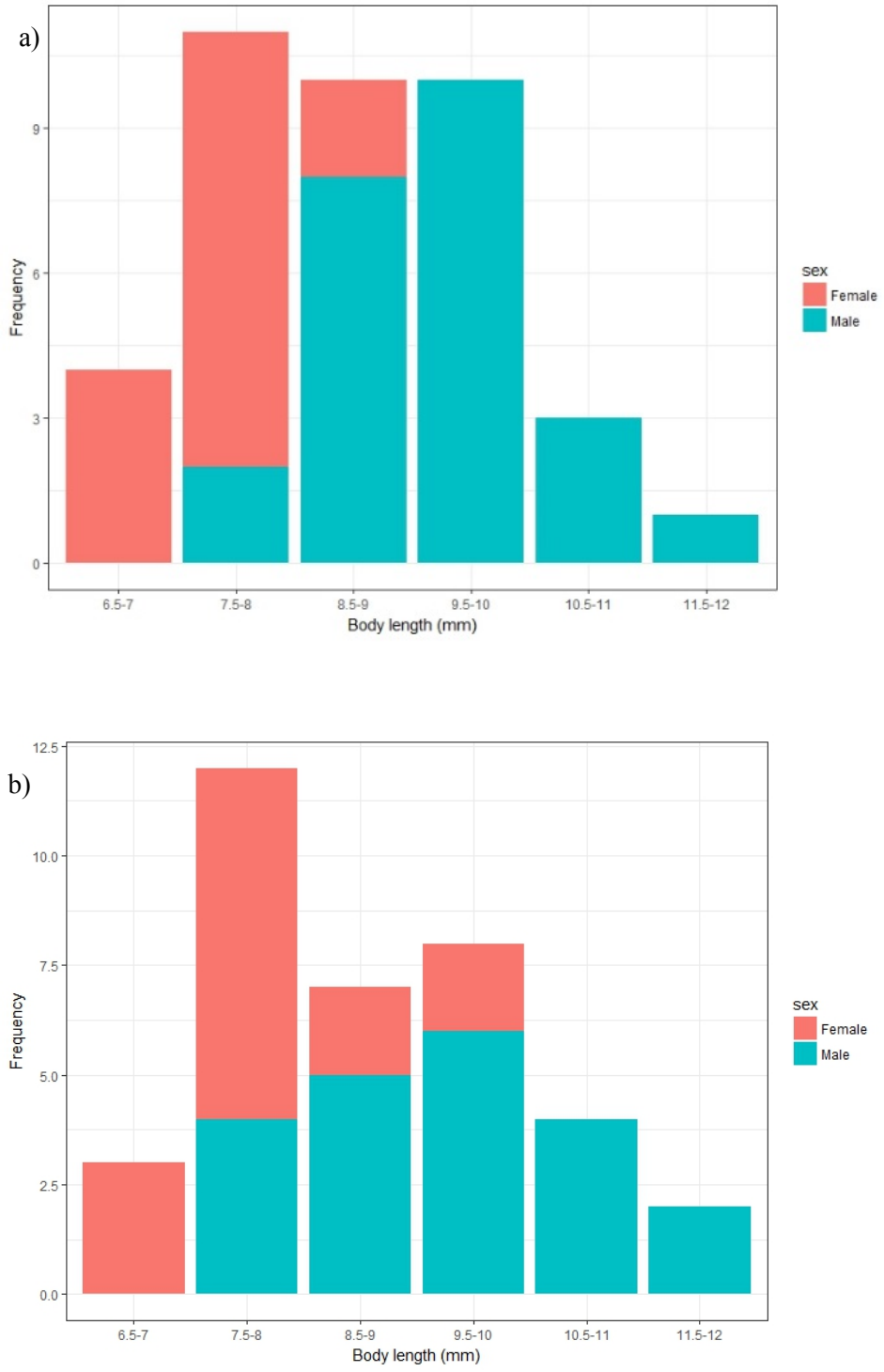


Figure S4.1: Correlation of the % *Gammarus* biomass in the sample, and the percentage *Gammarus* sequence reads (Pearson's Correlation Coefficient r = 0.747, df = 46,

P = 1.098 x 10^{-9}) Note that *G. fossarum* and *pulex* sequencing data have been combined here as the species were not distinguished during the initial morphological determination.

Figure S4.2. Frequency distribution of body length of male and female *Gammarus fossarum* individuals collected from a) the River Taff on 7/6/2016 and b) the River Frome on 27/6/2016



Chapter 5 Understanding the potential application of metabarcoding macroinvertebrate communities for ecological assessment in rivers: a cautionary tale

5.1 Abstract

Ecological assessment plays an important role in the conservation of freshwater habitats. Traditional methods that sample macroinvertebrate communities are well established and are commonly used to assess the impact of environmental pressures on water bodies, but can be time consuming and identification is often restricted by analyst skill, specimen quality and/or life stage. The rapid development of molecular tools for biodiversity assessment and the ability to detect rare, protected or invasive species without physical sorting and morphological identification of specimens opens an exciting new avenue of research and with it, potential applications for ecological assessment. Here we present the results of a large-scale comparison of traditional and environmental DNA (eDNA) sampling techniques along with microscopy-based morphotaxonomy and molecular characterisation of macroinvertebrate community composition. Four sample types were analysed from 44 sites in seven catchments in UK rivers: a 3-minute kick-sample, bulk DNA samples (organisms from the kick-sample), and eDNA from water and sediment. Our results show that all four sample types are significantly different in community composition with a variety of taxa unique to molecular methods including both target macroinvertebrate species used in EU Water Framework Directive monitoring and non-target species such as algae, diatoms and fish. However, the number of target macroinvertebrate sub-Operational Taxon Units, (sOTUs) was lower for molecular methods (bulk 68, water eDNA 16 and sediment eDNA 23), than for microscopy (157). Data collected from molecular methods produced lower status WFD classifications than those produced using traditional methods. This suggests a need to use the DNA derived data outputs to develop specific molecular based metrics for use in ecological assessment, particularly for the sOTUs detected via eDNA sampling, which do not form part of current directive reporting assessments. This study summarises the benefits and potential application of these new sampling methods, and highlights the importance of utilising this new tool in conjunction with traditional monitoring to inform development in new molecular driven ecological assessment tools.

5.2 Introduction

Freshwater macroinvertebrates are a key group of organisms used in biomonitoring programmes such as the European Union Water Framework Directive (WFD, 2000/60/EC) to classify the ecological status of river waterbodies. Macroinvertebrates form a major part of the biological

community in rivers and are included in biomonitoring programmes because they demonstrate a wide range of sensitivity to pollution types, exist in high numbers and have a relatively long lifespan, making them good indicators of environmental and anthropogenic pressures, particularly organic pollution (Metcalfe 1989; Bonada et al., 2006). The response of a macroinvertebrate community to environmental stress is characterised using metrics developed from known tolerances of different taxa particular pressures. The Whalley Hawkes Paisley Trigg (WHPT) metric, for example, is used in the River Invertebrate Classification Tool software (RICT) (Davy-Bower, 2007) to assess general water body degradation and predict and apply a classification to a river. RICT compares the macroinvertebrate community under conditions of no or minimal impact (reference site condition) with that observed at a site to quantify the impact of a pressure and classify the river accordingly (https://www.wfduk.org/resources/rivers-invertebrates-general-degradation).

Several established macroinvertebrate sampling methods are used in the UK under the WFD: sweep-netting, airlift and three-minute kick sampling, supplemented by hand searches (Environment Agency, 2017). These established sampling methods have constraints, with a three-minute kick sample only recovering approximately 62% of families and 50% of species within a given site (Furse et al., 1981). Increasing the sampling effort increases the number of specimens retrieved, however, due to mounting costs of sampling and sample processing times, alternative methods may be more efficient both in time and money spent for a similar level of ecological assessment. The characterisation of the community is based on morphology-based identification of specimens. This can be variable in terms of taxonomic resolution depending on the experience of the analyst and is prone to errors (Haase et al., 2006). Morphotaxonomy is particularly challenging for identifying cryptic species, juvenile stages of some species, and difficult groups such as Diptera. In addition, these established methods may not be the most efficient method for carrying out ecological assessment for crucial species, such as rare, protected and invasive taxa. Recent developments in the application of molecular tools for assessing biodiversity, such as metabarcoding of bulk samples and environmental DNA (eDNA), have the potential to enhance routine ecological monitoring and assessment.

High Throughput Sequencing (HTS) has been shown to be an effective tool for the detection of whole communities within freshwater, marine and terrestrial systems (Hajibabaei et al., 2011;

Shokralla et al., 2012; Carew et al., 2013; Hänfling et al., 2016; Elbrecht et al., 2017; Simmons et al., 2015; Brown et al., 2016). Metabarcoding of bulk samples, where DNA is extracted from whole communities of organisms, has demonstrated superior taxon identification in comparison to traditional methods, particularly for cryptic or hard to identify species such as Diptera (Sweeney et al., 2011; Yu et al., 2012; Jackson et al., 2014; Elbrecht et al., 2017; Blackman et al., 2017a). The largest comparison of routine macroinvertebrate monitoring via kick-sampling and metabarcoding of bulk samples, to date, was carried out by Elbrecht et al., (2017). They identified twice the number of taxa and found greater taxonomic resolution with metabarcoding than with morphotaxonomy (Elbrecht et al., 2017). Similar results for ecological status assessment were also found for the two methods (Elbrecht et al., 2017). However, issues with primer bias caused some taxa to be underrepresented or not amplified at all (e.g. Ephemeroptera: Baetidae, Elbrecht et al., 2017).

Although a bulk DNA sample approach for assessing the composition of aquatic macroinvertebrate communities has been demonstrated as a potential ecological assessment tool, the sample is still subject to the constraints of specimen collection in the field and processing and 'sorting' of specimens, which is both a costly and time consuming exercise (Elbrecht et al., 2015). A promising alternative is to sample DNA from environmental samples (i.e. environmental DNA or "eDNA") such as water, sediment or air (Valentini et al., 2009; Tablerlet et al., 2012; Rees et al., 2014; Bohmann et al., 2014; Lawson Handley 2015; Creer et al., 2016). Environmental DNA is a rapidly developing method for biodiversity assessment of communities (Hänfling et al., 2016; Port et al., 2016; Valentini et al., 2016; Deiner et al., 2017) and has been described as a 'game-changer' in terms of ecological monitoring (Sutherland et al., 2013; Lawson Handley 2015). eDNA taken from aquatic environments allows detection of species without collection (Ficetola et al., 2008) and is being increasingly used for community-level detection for example in mammals (Foote et al., 2012; Ushio et al., 2017), amphibians (Valentini et al., 2016), fish, (Kelly et al., 2014; Miya et al., 2015; Hanfling et al., 2016; Valentini et al., 2016; Port et al., 2016) and macroinvertebrates (Klymus et al., 2017; Macher and Leese, 2017).

eDNA not only allows the accurate identification of species within a site, but may also allow for further detection of those species which are often missed during physical collection methods, such as low abundant species including rare species or invasive non-native species (INNS). Sampling

and analysing eDNA may be seen as a means of bypassing lengthy and 'dirty' methods of analysing ecological samples, however, there are a number of important factors that influence DNA availability, which should be considered. For example, water samples within a river will be subject to several environmental variables, such as flow, temperature, UV, and pH, and their impact on DNA persistence along a water body is not yet fully understood (Shogren et al., 2017). Sediment eDNA is likely to retain DNA over longer timescales than water (Turner et al., 2015), and therefore may not represent true ecological status at the time of sampling. The surface of the substrate is also still exposed to the variables caused by flow such as DNA retention and resuspension (Shogren et al., 2017).

In order to fully understand the benefits and potential limitations of these new methods, a logical first step is to compare species detection using both DNA and morphology-based identification methods. In this study, we present the findings of the largest comparison of traditional kick-sampling and analysis with metabarcoding to date. The specific objective of the study was to compare the taxonomic information gathered from 4 sample types (kick-samples, bulk DNA, water eDNA and sediment eDNA) taken from 44 freshwater sites in the UK. We also carried out basic WFD classifications, based on presence/absence data to family level, as a first step in understanding how these methods may be used for ecological assessment.

5.3 Methods

5.3.1 Sampling and sample processing

Field surveys were carried out in March 2015 within 7 UK river catchments (Figure 5.1 and Supplementary Information I Table S5.1). Three sample types were collected at each of 44 sites: a three minute kick-sample (Murray-Bligh 1999) for mixed-taxon level identification by microscopy analysis and high molecular weight DNA extraction from pools of all specimens present in the kick-sample (here after referred to as bulk DNA sample), and a water and sediment sample for eDNA extraction (n = 176 in total). At each site environmental variables including water depth, width, substrate type and surrounding habitat were recorded. The method for sampling, filtration and extraction of the three molecular samples followed that of Blackman et al., (2017a). In brief; once identified by microscopy (full description in Macroinvertebrate microscopy analysis), each kick-sample was lysed in a Qiagen Tissue Lyser® and a 200 µl

subsample of the lysed tissue was then extracted using the DNeasy Blood & Tissue Kit® (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Two litre water samples were collected by pooling 4 x 500 ml sub samples from across the river width. Samples were then filtered through sterile 47 mm diameter 0.45 μm cellulose nitrate membrane filters and pads (Whatman, GE Healthcare, UK), using Nalgene filtration units (Thermo Fisher Scientific) in combination with a vacuum pump (15~20 in. Hg, Pall Corporation) in a dedicated eDNA laboratory at the University of Hull, UK. Sediment samples were collected in a similar way to the water samples, by collecting sub samples of the top layer of sediment from different points across the river width. Two-hundred millilitres of each sediment eDNA sample was placed in a sterile container with molecular grade water, then inverted 10 times and left to stand for 30s. The supernatant was then poured off into a sterile container, and 250 ml of this prefiltered through sterile 20 μm filter paper (Whatman, GE Healthcare, UK). The filtrate was subsequently filtered through 0.45 μm cellulose nitrate filters, as for the water samples. Environmental DNA from both water and sediment samples was extracted using PowerWater® DNA Isolation Kit (MoBio Laboratories, Inc. Carlsbad, USA) following the manufacturer's instructions. All sampling and filtration equipment was sterilized in 10% commercial bleach solution for 10 minutes then rinsed with 10% MicroSol detergent (Anachem, UK) and purified water between samples. Sample bottles filled with ddH_2O were taken into the field and filtered as sample blanks, 2L of ddH_2O filtration blanks were also filtered.

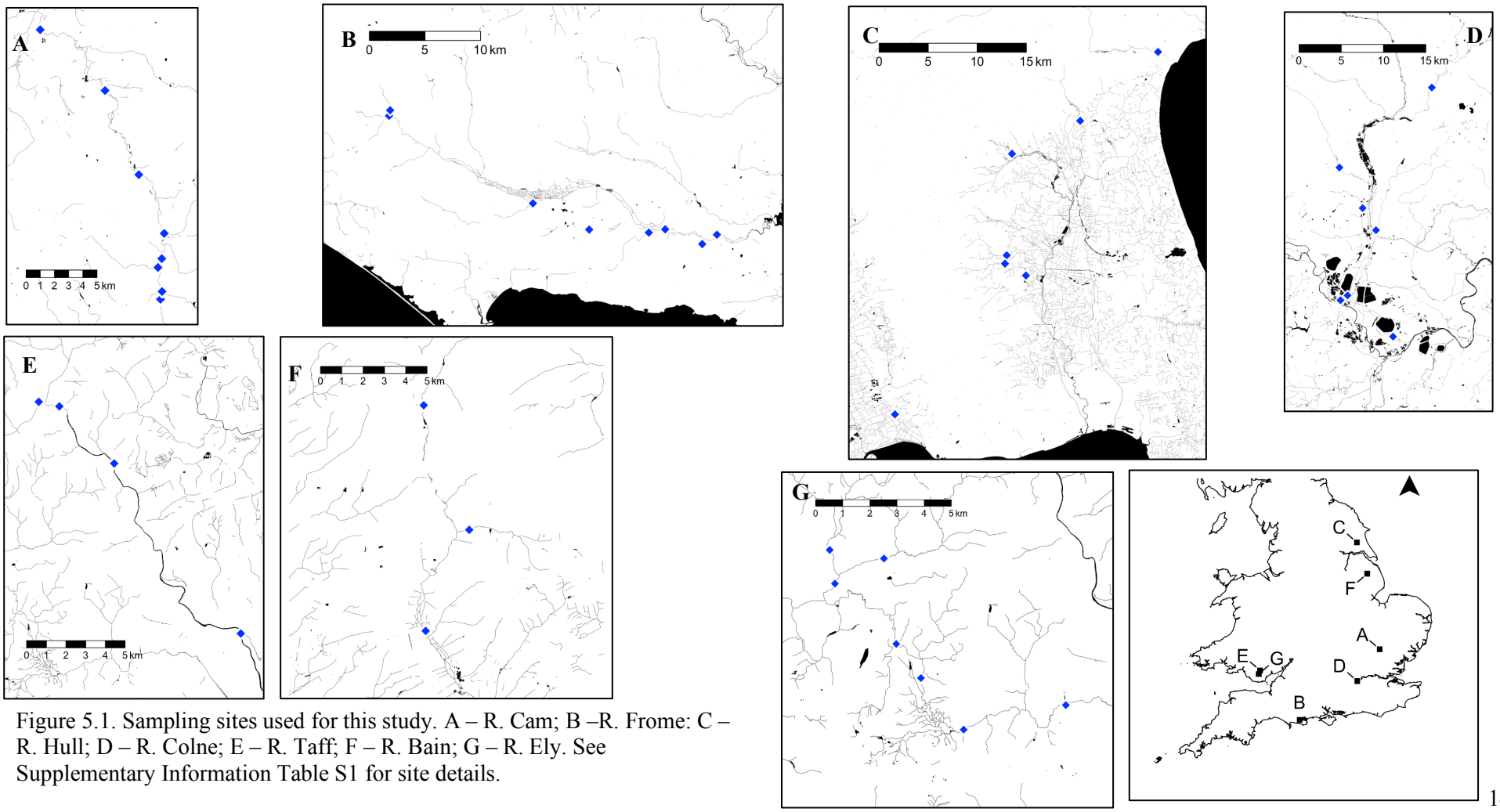


Figure 5.1. Sampling sites used for this study. A – R. Cam; B –R. Frome: C – R. Hull; D – R. Colne; E – R. Taff; F – R. Bain; G – R. Ely. See Supplementary Information Table S1 for site details.

5.3.2 Macroinvertebrate microscopy analysis

All macroinvertebrates from each kick sample were individually picked and identified to the River Invertebrate Prediction and Classification System (RIVPACS) taxon level 5 (mixed taxon). Mixed-taxon level analysis requires species determination of specimens wherever possible, with the exception of Oligochaeta, Chironomidae and other Diptera families for which Tribe/Genus level analysis is sufficient for WFD classification (See supplementary information I Table S5.2 for full mixed-taxon level guidance). Once morphological analysis was complete, specimens were stored in sterile falcon tubes filled with 100% ethanol. Total sample biomass and average species biomass was recorded for all samples.

5.3.3 PCR, library preparation and sequencing

The Cytochrome Oxidase subunit I (*COI*) gene was chosen for metabarcoding because it has the broadest taxonomic coverage for macroinvertebrates in public sequence databases and is the most widely used DNA barcode for taxonomic discrimination of macroinvertebrates. A 313 bp fragment was targeted using the primers described in Leray et al., (2013) which were designed to target invertebrate DNA. These primers were validated by carrying out in vitro tests on single macroinvertebrate species tissue samples to check for positive amplification. Library preparation followed a two-step process. First step PCRs were performed with primers jgHCO2198 TAIACYTCIGGRTGICCRAARAAYCA and mICOIintF GGWACWGGWTGAACWGTWT AYCCYCC (Leray et al., 2013). PCRs were carried out in 50 µl volumes containing 12.5 µl Taq DNA polymerase (Sigma Aldrich, UK), 0.2 µM of each primer, 1.5 mM $MgCl_2$, 2 µM dNTPs, and 2 µl of undiluted DNA template. PCRs were performed on an Eppendorf Mastercycler using the following conditions: initial denaturation 95°C for 2 mins, followed by 35 cycles of denaturation 95°C for 15s, annealing at 50°C for 30s and extension at 72°C for 30s, with a final extension time of 10 min at 72°C. Samples included PCR and filtration blanks (n=3 and n=5, respectively) as negative controls and single species positive controls (*Triops cancriformis* (Bosc, 1801) (n=2) and *Osmia bicornis* (Linnaeus, 1758) (n=2)). PCR products were confirmed by gel electrophoresis on a 1.5% agarose gel stained with SYBR Safe (Invitrogen, UK). Post first PCR clean-up was carried out using ZR-96 DNA Clean-up Kit (Zymo) following the manufacturer's protocol.

Individual Nextera tags (Ilumina) were added in a second PCR using 0.25 μM of each tagging primer and 1 μl of purified PCR product. PCR conditions were: initial denaturation at 95°C for 2 min, followed by 8 cycles of denaturation at 95°C for 15s, annealing at 55°C for 30s and extension at 72°C for 30s, with a final extension time of 10 mins at 72°C. Samples were normalised using SequalPrep Normalization plates (Invitrogen, UK) including positive and negative controls (including sample and filtration blanks). Each plate of samples was pooled, quantified using the Qubit dsDNA HS kit on a Qubit fluorimeter (Thermo Fisher Scientific, UK) and each plate combined in equimolar quantities to make the final library. The pooled library was concentrated using a centrifugal vacuum concentrator (Eppendorf, UK), and checked on a 2.0% agarose gel. The band containing the DNA was extracted and cleaned using QIAquick Gel Extraction Kit (Qiagen) following the recommended manufacturer's protocol. The library was run at 6 pM concentration on an Illumina MiSeq, with 20% PhiX using the 2 x 250bp V2 chemistry.

5.3.4 Bioinformatics

The bioinformatics pipeline for post-sequencing involved the following steps: sequenced paired-end reads were joined using PEAR (Zhang et al., 2014), quality filtered using FASTX tools (Hannon, http://hannonlab.cshl.edu) and chimeras were identified and removed with VSEARCH_UCHIME_DENOVO (Rognes et al., 2016). The sequences from the samples were demultiplexed, quality checked and clustered sub-Operational Taxonomic Units (sOTU) using DEBLUR (Amir et al., 2017). A COI database was created (see below) against which sOTU representative sequences were validated by BLASTN analysis. For each sOTU, BLASTN results were screened and hits with the following parameters were retained: maximum e-value of 1e-10, minimum bit-score of 175 and maximum number of hits of 20. Where there was a hit or hits above a threshold of 97% sequence identity, the taxonomic classification of the hit with the highest sequence identity was assigned to the sOTU. For those without hits above 97% sequence identity, the lowest common ancestor of the taxonomic classifications from the hits was selected as its taxonomic assignment. We quantified the level of possible contamination using sequences detected in the Triop cancriformis and Osmia bicornis single species positives, which enabled us to choose a suitable threshold level (0.5% of total sample reads) for filtering and removal of low level contamination, in line with similar studies (Port et al., 2016; Hänfling et al., 2016; Blackman et al., 2017a).

5.3.5 COI database

Sequences were retrieved from the NCBI nucleotide database using Entrez Programming Utilities (http://www.ncbi.nlm.nih.gov/books/NBK25501) using search terms "COI OR cox1 NOT uncultured NOT environmental", ensuring sequences of all taxa identified during microscopy were present in the database. From this dataset, sequences with length outside 500 and 2000 bps were removed and redundant sequences were filtered out using VSEARCH (github.com/torognes/vsearch). The remaining sequences were translated to six reading frames by a custom script utilising Biopython's Bio.SeqUtils library (Cock et al., 2009). In order to encompass homologues of COI in bacteria and archaea, all publically available genomes of bacterial and archaeal origin were downloaded from NCBI Genomes database (https://www.ncbi.nlm.nih.gov/genome/) and their open reading frames were subsequently retrieved using FragGeneScan (Rho et al., 2010). Both sets of translated sequences were combined, and to check for gene content the sequences were structurally assessed for its COI-intactness using HMMER3 (Eddy., 1998) against a Hidden-Markov models profile derived from a COI pfam seed alignment (Finn et al., 2014). These structurally verified COI sequences were then taxonomically assigned with NCBI Taxonomy based on their gene identifier.

5.3.6 Data Analysis

All statistical analyses were done with R 3.4.2 (R-Core-Team 2017) using the packages Vegan (Oksanen et al., 2017), ggplot2 (Wickham et al., 2009) and Venn.Diagram (Chen, 2016). Non-metric multidimensional scaling (NMDS) was used to visualise similarities between sample types, and statistical differences between groups was tested using Analysis of Similarity (ANOSIM) (Zuur et al., 2007). In order to compare WFD river classifications from outputs using both metabarcoding and traditional microscopy, 5 river sites with good historical and environmental data were used. The River Invertebrate Classification Tool (RICT) was used to classify the ecological quality of rivers (Davy-Bowker et al., 2007). Following guidance from UKTAG River Assessment Method Benthic Invertebrate Fauna (www.wfduk.org) only presence/absence data of target macroinvertebrate Families level data (as opposed to abundances) were used to calculate both WHPT number of taxa (N-taxa) and WHPT Average Score Per Taxon (ASPT) indices. The data derived from our four sample types was compared to the expected values from the RICT software. This was then used to calculate Ecological Quality Ratios (EQR) and compared to WFD classification boundaries, and we used the lowest scoring EQR to classify the site. It must be

noted that these classifications are based on data from Spring sampling only, in reality a classification would be based on data from a full year with both Spring and Autumn samples (https://www.wfduk.org/resources/rivers-invertebrates-general-degradation).

5.4 Results

5.4.1 Kick samples

In this section, we will refer to all freshwater macroinvertebrate sOTUs used for WFD classification as "target-taxa" (See supplementary information Table S5.1 for the list of taxa identified to mixed-taxon level and used for WFD classification). All other sOTUs will be referred to as "non-target taxa". We did not include any sOTUs identified to family level or lower taxonomic level which were not solely aquatic taxa in the "target-taxa" category, however we have included all reads identified to Eukaryota level only, in order to represent the data accurately. All in vitro single species tests amplified using the primer set (See Supplementary Information I Table S5.3 for details of species tested) and sequences for all taxa identified via microscopy were in the database.

The total sequence read count passing quality control, after removal of chimeric sequences, across all three sample types was 2,966,758. After applying a 0.5% threshold to the samples, the total read count was 2,699,197 with a mean read count of 18744 per sample. We detected 247 operational taxonomic units (sOTUs) in total, including a large number of non-target sOTUs from fish, algae and bacteria. Of the 44 sites, only 30 sediment samples and 15 water samples yielded target taxa reads. Figure 5.2A shows a summary of the read numbers found for all taxa in each sample type while Table 5.1 summarises the mean read count and total sOTUs found in each of the molecular methods for all species and then target taxa only. A large proportion of the reads from water and sediment samples (33.42% and 37.87% respectively) could only be classified to Eukaryota. The large number of non-target taxa in the sediment and water samples is likely due to reference sequences being unavailable for species found in these sample types, and compares to only 0.5% of bulk DNA sample reads. Water eDNA samples contained 35 bacteria, algae and rotifer sOTUs, many of which were identified to species level (Fig. 5.2A). The non-target invertebrate taxa from the water eDNA came from a range of sources, including both aquatic and

non-aquatic sOTUs. Overall across the sample types, 87.88% of reads from bulk DNA, 24.73% of the reads from water and 6.52% of the reads from sediment were from target taxa (Figure 5.2B).

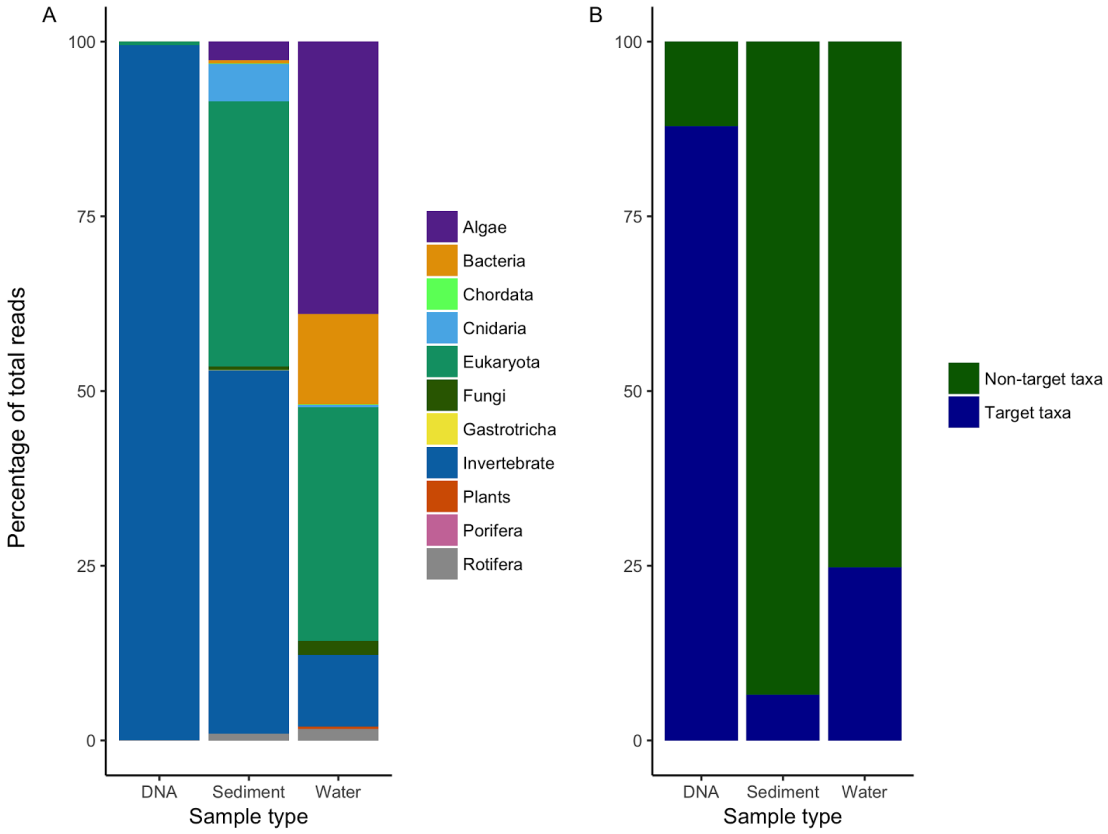


Figure 5.2: Summary of read counts from each molecular sampling method. A. Summary of the percentage of each group/order present in the samples before non-target sOTUs were filtered out (the Algae group includes Diatom taxa). B. Target and non-target invertebrate sOTUs detected in the three sample types.

Table 5.1: Summary of read count and sOTUs. All taxa reported from the three molecular sample types collected, Microscopy total target taxa only has also been included.

Sample type	All data		Target taxa only	
	Average No. of reads number per sample	Total number of sOTUs	Average No. of reads number per sample	Total number of target sOTUs
DNA – bulk	27045	108	23638	67
Water eDNA	10714	101	273	16
Sediment eDNA	16813	123	570	23
Microscopy	-	-	-	156

5.4.2 Comparison of sample types

When all taxa were included (target and non-target) there was a highly significant difference in community composition between sample types (ANOSIM, R: 0.7737, P = 0.001) (Fig. 5.4A), and between each paired sample type (Table 5.2, Blue). Only 79 of the 247 sOTUs detected using metabarcoding were from target taxa, compared to 156 target taxa identified by microscopy. Coverage of the different Class/Orders made by each of the methods show the limited detection by metabarcoding (Fig. 5.3), with no taxa recorded in Hemiptera, Megaloptera or Plecoptera, despite successful amplification of Megaloptera and Plecoptera in in vitro testing.

After the removal of non-target taxa reads, of the 44 sites surveyed, only 11 sites had target taxa reads from all four sample types due to the high proportion of non-target taxa amplification. We therefore carried out a target taxa based NMDS for these 11 sites only (Fig. 5.4B). Although there is some overlap between sample types, globally they were significantly different (ANOSIM R:0.6291, P = 0.001). ANOSIM analysis between paired sample types also show a number of sample types are significantly dissimilar (Table 5.2 green). However, the dissimilarity between the molecular sample types: sediment and DNA, DNA and water, sediment and water are not significant (R: 0.07558 P = 0.103, R: 0.1283 P = 0.03, R: 0.04313 P = 0.131, respectively, Table

5.2, green). This information is also shown in the Venn diagram (Fig. 5.4C) to the highest taxonomic resolution the sequence or specimen was reported to. As displayed in the figure the number of sOTUs present in only microscopy is the largest, 123 and only 3 taxa were found in all three methods (*Asellus aquaticus*, Chironomidae and *Ephemera danica*). However, there are a total of 46 sOTUs found from the three molecular methods and not in microscopy, including 28 found in the bulk DNA samples which could be potential gut content and, in some cases, greater taxonomic identification (See Supplementary Information II Table S5.4 for the full Venn diagram output).

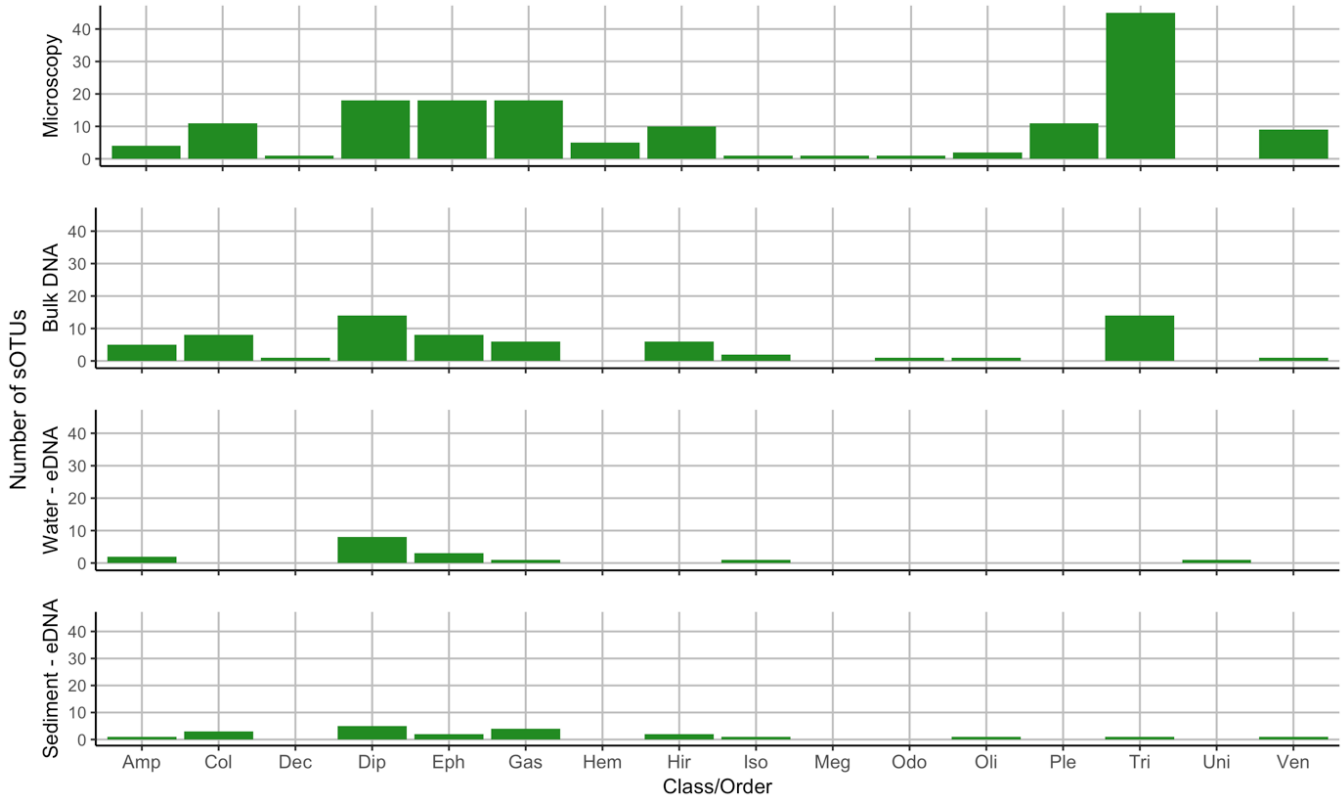
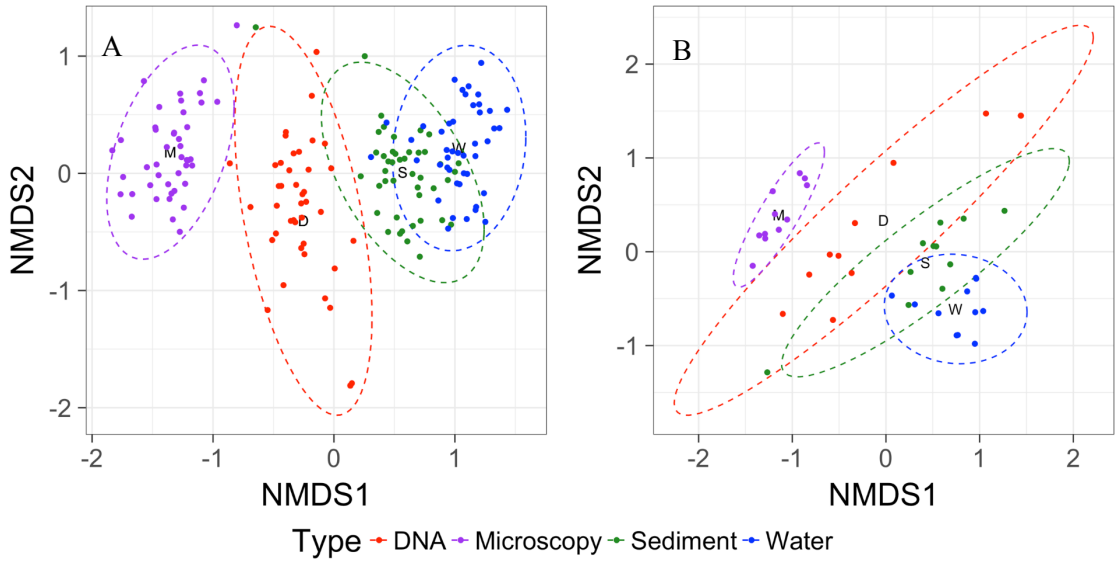


Figure 5.3: sOTUs to Class/Order level for all four sampling methods. This summary shows the number of families recorded in each Class/Order group gained from each of the sampling methods with all samples combined. Amp – Amphipoda, Col - Coleoptera, Dec – Decapoda, Dip - Diptera, Eph - Ephemeroptera, Gas Gastropoda, Hem - Hemiptera, Hir - Hirunida, Iso - Isopoda, Meg - Megaloptera, Odo - Odonata, Oli - Oligochaeta, Ple - Plecoptera, Tri - Trichoptera, Uni - Unionida, Ven - Veneroida.



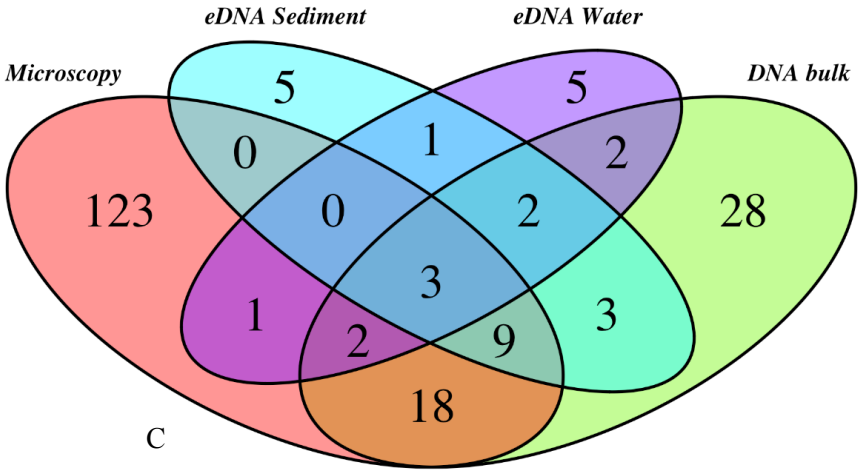


Figure 5.4: Comparison of methods. A and B Non-multidimensional scaling (NMDS) for all four sample types for all taxa (A) and for target taxa only (B). Points represent the samples from each sample type, stress value 0.10318 (A) and 0.09277581 (B). M – microscopy, D - DNA bulk, S – sediment eDNA and W – water eDNA. C. Sample type Venn diagram comparison of target taxa detected. This diagram shows the sOTUs found in each sample type separately and with other types. (See Supplementary information II for Venn diagram output summary)

Table 5.2: ANOSIM results from paired sample type analysis. Blue - ANOSIM results based on all (target and non-target) sOTUs from all sample types. Green - ANOSIM results based on target taxa only for all sample types

Sample type	DNA bulk	Sediment eDNA	Water eDNA	Microscopy
DNA bulk	-	*R: 0.07558* *P = 0.103*	*R: 0.1283* *P = 0.03*	R: 0.4132 P = 0.001
Sediment eDNA	R: 0.5912 P = 0.001	-	*R: 0.04313* *P = 0.131*	R: 0.429 P = 0.001
Water eDNA	R: 0.7181 P= 0.001	R: 0.249 P = 0.001	-	R: 0.4211 P = 0.001
Microscopy	R: 0.7221 P = 0.001	R: 0.9656 P = 0.001	R: 0.9972 P = 0.001	-

5.4.3 Water Framework Directive river site classification

WFD classification was derived from the lowest of the two EQRs (N-taxa or ASPT), in all cases the DNA-based sample types are classified using the N-taxa. This is a result of the low number of target taxa recorded by these methods, particularly with regards the eDNA samples with EQR values ranging from 0.07 - Poor status to 0.31 - Bad status. However, in all but one site (Water Lane, Newport), the bulk DNA sample ASPT score is much closer to the expected classification score, with EQR values ranging from 0.91 - Good status to 1.07 - High status (Table 5.3).

Table 5.3: Water Framework Directive classification. This table summarises the classification for number of taxa (N-taxa) and average score per taxa (ASPT) based on the Whalley, Hawkes, Paisley & Trigg (WHPT) for all sample types at 5 sites selected from our dataset. The scores have been calculated using RICT and overall classification derived from using the "worst of" approach as stated in the UKTAG guidelines.

Site	Sample type	Observed N-taxa	Expected N-taxa	EQR	Classification	Observed ASPT	Expected ASPT	EQR	Classification	Overall classification	Actual classification
B1383 Road Bridge, Upper Cam	Microscopy	18	26.087	0.67	Good	5.8	5.36	1.08	High	Good	High
	DNA – Bulk	8	26.087	0.31	Bad	5.5	5.36	1.07	High	Bad	
	Water – eDNA	5	26.087	0.19	Bad	4.8	5.36	0.89	Good	Bad	
	Sediment - eDNA	3	26.087	0.11	Bad	4.6	5.36	0.85	Moderate	Bad	
Water Lane, Newport, Upper Cam	Microscopy	16	25.782	0.62	Moderate	4.8	5.35	0.89	Good	Moderate	High
	DNA – Bulk	2	25.782	0.08	Bad	3.6	5.35	0.67	Poor	Bad	
	Water – eDNA	8	25.782	0.31	Bad	4.0	5.35	0.74	Moderate	Bad	
	Sediment - eDNA	3	25.782	0.12	Bad	2.7	5.35	0.50	Bad	Bad	

Hauxton Mill, Upper Cam	Microscopy	16	26.487	0.60	Moderate	5.5	5.43	1.02	High	Moderate	High
	DNA – Bulk	9	26.487	0.34	Bad	5.4	5.43	0.99	High	Bad	
	Water – eDNA	4	26.487	0.15	Bad	3.1	5.43	0.57	Poor	Bad	
	Sediment - eDNA	4	26.487	0.15	Bad	5.2	5.43	0.94	Good	Bad	
US Colne Brook, River Wraysbury	Microscopy	16	27.517	0.58	Moderate	5.2	5.524	0.93	Good	Moderate	Good
	DNA – Bulk	7	27.517	0.25	Bad	5.1	5.524	0.91	Good	Bad	
	Water – eDNA	2	27.517	0.07	Bad	3.2	5.524	0.58	Poor	Bad	
	Sediment – eDNA	4	27.517	0.14	Bad	4.6	5.524	0.83	Moderate	Bad	
Ashford Common, River Wraysbury	Microscopy	18	27.396	0.66	Moderate	4.1	5.415	0.75	Moderate	Moderate	Good
	DNA – Bulk	3	27.396	0.11	Bad	5.1	5.415	0.94	Good	Bad	
	Water – eDNA	3	27.396	0.11	Bad	2.9	5.415	0.53	Bad	Bad	
	Sediment - eDNA	3	27.396	0.11	Bad	4.9	5.415	0.90	Good	Bad	

5.5 Discussion

In this study, we compared metabarcoding of different sample types (bulk DNA and eDNA samples from sediment and water) with traditional morphology-based identification of kick-samples, to determine the macroinvertebrate community at 44 river sites across 7 catchments in the UK. The number of target macroinvertebrate taxa collected by kick-samples was higher than the number of target macroinvertebrate sOTUs generated from all three DNA sample types combined. The current WFD classification methods are unsuitable for accurately describing the extra information gained from metabarcoding data, and therefore there is currently a significant shortfall in metabarcoding to accurately classify sites. However, we see a large number of non-target taxa, identified to species level, in the molecular data which could be used to inform classification of waterbodies in the future.

5.5.1 Comparison of microscopy and metabarcoding

The first aim of this study was to investigate the comparability of taxa (sOTUs) retrieved from three different DNA sample types with traditional morphology-based detection from kick-samples. The metabarcoding data detected 247 sOTUs across the three different sampling methods, including a large proportion of non-target aquatic and terrestrial sOTUs. This compares to 156 target taxa identified using kick sampling. Of the DNA based sOTUs, 79 were freshwater macroinvertebrates. Twelve of the 15 major macroinvertebrate groups (Classes/Orders) identified in the microscopy were detected by metabarcoding the bulk DNA. The bulk DNA sample data revealed that a large proportion of taxa were not detected by metabarcoding but found via microscopy (124 sOTUs). Some of these taxa may have been missed due to differing levels of taxonomic identification by the two methods.

As this is a developing field, there are several elements still under consideration, which will have influenced the detection of taxa in this study, particularly availability of sequences in reference databases, primer bias, and bias due to differences in biomass between taxa. Firstly, reference database coverage is a limiting factor in metabarcoding studies particularly of poorly studied taxa and habitats. In this study. we ensured the taxa detected by microscopy were in the reference database, however species which may have been detected by metabarcoding but not identified

during microscopy (e.g species detected only by molecular methods) may not have been represented and therefore be missing from the data, furthermore although representatives of each taxa were present, haplotype diversity may not have been represented in full. Both the water and sediment eDNA samples contained a large proportion of reads identified to Eurkaryote only. A likely source of this classification is the lack of reference sequences found in current databases and highlights the need to generate reference sequences for accurate identification of metabarcoding sequences. Similarly, studies such as this face restrictive primer choice due to reference database availability.

Secondly, three taxonomic groups are missing in the metabarcoding data, despite being recovered via microscopy and two of these groups (Plecoptera and Megaloptera) successfully amplifying during in vitro testing. This could result from primer bias in a mixed sample and is a key concern when using universal primers (Elbrecht et al., 2015; Elbrecht et el., 2016). Here, the COI region was chosen as sequences are widely available for most target macroinvertebrate species and has been frequently demonstrated to be the preferred region for macroinvertebrate studies (Tang et al., 2012; Elbrecht et al., 2017), our choice of universal primers therefore was restricted and often these primers have been developed for a specific application. The primers used in this experiment aim to have complete metazoan coverage and were designed for fish gut content in a marine environment (Leray et al., 2013). Although suitable for other applications they may show preferential amplification of certain taxa over others. Further testing of these primers by the authors showed up to 35% of taxa within mock communities were undetected (Leray and Knowlton, 2015, Elbrecht et al 2017) and this is seen in our study with three missing groups (Hemiptera, Plecoptera and Megaloptera). The Leray (2013) primers have also been shown to produce fewer sequences and have poor detection of invertebrate taxa compared to other universal primers (Alberdi et al., 2018) and therefore a "multi-marker" approach may be more appropriate for these sampling methods (Deagle et al., 2014; Alberdi et al., 2018), which will provide information from a range of different taxa groups from single samples using the metabarcoding approach and may go further in addressing the issue of primer bias.

Thirdly, a large proportion of Family level taxa are missing completely in the bulk DNA sample data. This could be due to biomass bias, as we did not size sort or discriminate between any taxon in our sample processing, however the species which were missed by molecular methods range

significantly in average biomass from 0.0008g (*Caenis luctuosa*) to 1.3g (*Dreissena rostriformis bugensis*). A previous study showed specimens of lower biomass will be missed in comparison to those with a higher biomass (Elbrecht et al., 2015) which is often the case here, but so too are some larger specimens. However, implementing a size sorting step for this form of sample process will remove much of the time advantages of metabarcoding and is not realistic for routine sample processing.

Although issues of missed groups and families are prevalent in this dataset, we do see higher taxonomic resolution in some cryptic groups, particularly Diptera consistent with other metabarcoding studies (Elbrecht et al., 2017). In comparison to traditional microscopy which recorded all Chironomidae and Simuliidae to Family, Genus or tribe level, the metabarcoding data confirmed 6 species within each family. Chironomidae in particular are noted as an important bioindicator (Bista et al., 2017) requiring highly skilled taxonomists to carry out full species level identification which is both time consuming and challenging. Here it has been demonstrated that metabarcoding data already yields far higher numbers of sOTUs than traditional kick-sampling, and though a large proportion of these sOTUs were not targeted by this study, this new information offers the opportunity to gain further ecological assessment of waterbodies using different groups which have previously been discounted. To this end we can see the taxa difference between sample types is highly significant both at target taxa and at all sOTU levels. This is clearly illustrated by the number and diversity of macroinvertebrate target taxa picked up by traditional kick-sampling and microscopy compared to metabarcoding. However, when we consider the number of non-target taxa removed from these figures it is impossible to carry out an accurate comparison without dismissing the true potential of metabarcoding information.

5.5.2 Comparison of molecular methods

A large proportion of the sOTUs detected from eDNA samples were from non-target taxa. Bulk DNA samples consisted mainly of target species and had the highest number of sOTUs of the molecular methods (68). Of the eDNA samples, we see a much greater number of non-target taxa detected. However, with regards the water samples, this was not unexpected as similar studies have also found eDNA samples from water to be dominated by algae, bacteria and rotifers (Deiner et al., 2016). Since much of the water eDNA sOTUs were algal species, it is intuitive to find alternative analysis methods in which this data could be used. Research by Wurzbacher et al.,

(2017) have demonstrated that metabarcoding entire groups of phytoplankton, zooplankton and bacterioplankton offers a new opportunity for greater ecological assessment based on multiple groups, rather than a single group. Similarly, of the sediment sOTUs, although a large proportion of the reads detected were from macroinvertebrates they are not recognised target taxa for WFD classification, resulting in a significant loss of data for ecological assessment metrics. A large number of the invertebrate sOTUs from sediment eDNA samples were from the Orders Haplotaxida and Lumbriculida, which are often disregarded and grouped as Oligochaeta by current WFD monitoring methods, as they are notoriously difficult to identify to species level due to cryptic diversity and a lack of distinctive features. However, this group is highly sensitive to environmental change and also a bioindicator. A recent study found a much higher number of species using molecular methods than traditional microscopy (Vivien et al., 2015), making this group an ideal candidate for metabarcoding identification. Using eDNA from river bed sediment may offer an exciting new resource in ecological assessment which is currently overlooked.

The key issues associated with detecting eDNA in water and sediment samples are production of DNA by target taxa and effects of flow and substrate type on DNA transport and retention. Firstly, considering DNA production, the amount of DNA produced by the target taxa will influence our ability to detect it. It is highly likely that a high proportion of the phytoplankton detected was in fact organisms rather than eDNA in its purest form, and the much lower detection of macroinvertebrate eDNA in our samples is down to varying DNA production rates by the different taxa. For example, previous studies have shown that mussels will produce DNA freely and have been regularly detected by other eDNA methods (Blackman et al, 2017c *in review*), in this study we found eDNA water samples were the only sample type to detect the presence of Unionidae. However, we also know other macroinvertebrate DNA production is sporadic and is often hard to detect in low densities (for example Crustacea: Treguier et al., 2014, Dunn et al., 2017) and therefore, along with other factors will contribute to a lack of detection in flowing water.

Secondly, lotic networks act as a "conveyor belt" for biodiversity information (Deiner et al., 2016) and it can be assumed that many of the sOTUs recorded in the water samples are from upstream points in the catchment rather than representative of the taxa at the fixed point of sampling. Currently, the general understanding of the effects of flow, substrate and other biotic factors on eDNA persistence is low and further investigation is needed (Shogren et al., 2017). Several studies

have established that DNA can persist within lotic waterbodies, with results ranging from meters to kilometres (Deiner and Altermatt, 2014; Jane et al., 2015). However, this is both a positive and negative feature of this sample type; as samples may reflect the wider catchment and therefore be useful to detect rare or new invasive species and inform ecological assessment on a river or catchment scale (Deiner et al., 2016). Similarly, as Macher & Leese (2017) showed, the eDNA signal differs not only along a river stretch but within a site, thus making it difficult to determine to what geographical extent the eDNA signal detected in this study represents.

Finally, several studies have shown substrate types have an effect on the retention and resuspension of DNA through lotic systems in particular finer substrates are able to retain eDNA quicker and for longer periods of time (Jerde et al., 2011; Shogren et al., 2017), this is also influenced by slower flows associated with finer sediments. However, it is still unclear as to how much or how little this effect will have on the overall DNA content of a sediment sample. Natural river systems, like the ones studied here have a variety of substrate types available, we therefore are unable to determine to what extent an eDNA sample may represent a contemporary biodiversity assessment or a much longer timescale (Turner et al., 2015).

5.5.3 Towards ecological classification with metabarcoding: how can we move forward?

As part of this study we aimed to demonstrate the current classification methods for monitoring macroinvertebrates using metabarcoding data, in line with the work by Hanfling et al., (2016) who had demonstrated metabarcoding data reflected the ecological status of lake basins on Lake Windermere. Although the results in this study do not represent a full classification in line with current WFD methodology (Spring and Autumn sampling), the data aimed to demonstrate the extent to which WFD classification may be inferred using metabarcoding. Contrary to Elbrecht et al., (2017), in our study we do not see comparable classifications of metabarcoding bulk DNA samples and kick-samples. It is clear that numerous groups of taxa have been missed causing a significantly low N-taxa score for all DNA-based sample types. However, there is some comparability in ASPT with relatively similar scores between microscopy and bulk DNA samples. As mentioned previously this should not in any way detract from the use of metabarcoding of bulk or eDNA sample, but perhaps highlight the need to adapt current assessment methods, or develop new ones using information derived from metabarcoding, allowing the incorporation of multiple species groups (macroinvertebrate, algae and fish) to be

used simultaneously in ecological assessments. The ecological assessment of a habitat is fundamentally constrained by the sampling method being employed and therefore, as demonstrated by the data presented here and by others (Deiner et al., 2016, Wurzbacher et al., 2017), we must acknowledge that metabarcoding allows a more holistic approach to biodiversity monitoring. Rather than separating the sample types, they should be considered in conjunction with one another to gain a fuller and more accurate picture of all taxa in the habitat being surveyed; algae, macroinvertebrates and fish, which will allow far greater biodiversity assessment.

5.6 Conclusion

The molecular methods described in this study provide a platform for further and greater ecological assessment. Metabarcoding may not perform in the same way as current morphological-based assessment so cannot currently be used to retrofit or replace existing methods. However, they provide an exciting opportunity for greater taxonomic identification (cryptic species) and have the ability to detect a combination of taxa across groups, some of which are not utilised in current ecological assessment methods but which have the potential to respond to environmental pressure gradients. Molecular methods are highly sought after as a method for complete bioassessment of a waterbody, and we believe this is achievable with further research, we further recommend the use of a multi-assay approach in order to investigate the full potential of information gained from eDNA samples and emphasize the importance of investing in comprehensive reference databases in order to maximize taxa detection.

5.7 Supplementary Information I

Table S5.1: Site details

Catchment	Site name	Latitude	Longitude
Upper Cam	GT CHESTERFORD ROAD BRIDGE	52.062132	0.19238021
Upper Cam	SLADE FARM BRIDGE	52.024818	0.21862944
Upper Cam	WHITTLESFORD	52.115541	0.15753502
Upper Cam	WENDONS AMBO	52.008769	0.21611596
Upper Cam	B1383 ROAD BRIDGE	52.003169	0.2121386
Upper Cam	ROAD BRIDGE DEBDEN WATER, NEWPORT	51.982932	0.21476742
Upper Cam	WATER LANE, NEWPORT	51.987856	0.21639406
Upper Cam	HAUXTON MILL	52.154102	0.090515423
River Hull	COTTAGE PASTURE - BRYAN MILLS BECK	53.903298	-0.45552486
River Hull	ARRAM CATCHWATER DRAIN BY STATION	53.884651	-0.42544783
River Hull	ELMSWELL BECK - SHOWGROUND	53.996909	-0.44733273
River Hull	GYPSEY RACE - BRIDLINGTON	54.090584	-0.21865512
River Hull	KELK BECK - U/S RAILWAY BRIDGE (HU17A)	54.027221	-0.34051031
River Hull	MILL (MIRES) BECK - ELLERKER	53.756246	-0.63056607
River Hull	SCORBOROUGH BECK @ SCORBOROUGH	53.895594	-0.45827442
Taff	At Blackweir	51.494675	-3.1960033
Taff	AT Upper boat	51.572643	-3.2893901
Taff	DS Pontypridd	51.59878	-3.3298356
Taff	US Taff	51.600856	-3.3449849
Ely	AT ST FAGANS	51.484711	-3.2695964
Ely	AT PETERSTON-SUPER-ELY	51.476645	-3.3240909
Ely	NEAR BRYN FARM	51.49356	-3.3469089
Ely	AT PONT TAL-Y-BONT	51.504738	-3.3600621
Ely	AT PONT Y CLUN	51.524496	-3.3927959
Ely	AT CROSS INN	51.532716	-3.366663
Ely	US L'OREAL	51.535523	-3.3955837
Bain	BISCATHORPE FORD	53.346675	-0.15228854
Bain	HEMINGBY	53.251269	-0.15088622
Bain	GOULCEBY	53.294029	-0.12005058
River Frome	Moigne Combe	50.684015	-2.3198474
River Frome	Seven Stars (12**)	50.681584	-2.2432145
River Frome	Wool Bridge	50.684243	-2.2221799

River Frome	LUCKFORD 10M D/S BRIDGE (12**)	50.67222	-2.1745039
River Frome	Holme Bridge	50.679854	-2.1556372
River Frome	West Stafford Rectory	50.705377	-2.3922157
River Frome	U/S Maiden Newton	50.776549	-2.5771882
River Frome	U/S Maiden Newton	50.781052	-2.5758252
COLNE (THAMES)	ABOVE COLNE BROOK	51.541392	-0.49943343
COLNE (THAMES)	AT BUSHEY MILL LANE, WATFORD	51.670499	-0.38036323
COLNE (THAMES)	ABOVE COLNE BROOK	51.442446	-0.53751562
COLNE (THAMES)	U/S GERARDS CROSS STW	51.584912	-0.53929734
COLNE (THAMES)	AT STAINES MOOR/HYTHE END	51.44766	-0.52523205
THAMES SOURCE TO TEDDINGTON (NE SECTION)	D/S ASHFORD COMMON WTW	51.403335	-0.44708218
COLNE (THAMES)	U/S FRAYS	51.517697	-0.47643214

Table S5.2: UK Environment Agency macroinvertebrate mixed-taxon level analysis guidance. The table gives the taxonomic level required during kick-net sample analysis required for each Phylum.

Taxonomic Group				Identification
Phylum	Class	Order/subclass	Family	
Platyelminthes	Turbellaria	Tricladida		All species aprat from species couplets, *Polycelis nigra/tenuis* and *Dugesia lugubris/polychroa*
Nematoda				Phylum
Nematomorpha				Phylum
Annelida	Oligochaeta			Class
	Hirudinea			Species
Mollusca	Mollusca			Species (except for *Pisidium* sp)
Arthropoda	Crustacea	Malacostraca		All adults to species and larvae to species or genus where keys allow
	Insecta	Ephemeroptera		
		Plectoptera		
		Odonata		
		Hemiptera		
		Coleoptera		
		Megaloptera		
		Trichoptera		
		Diptera		Family except the specified below and cranefly genera/species Tipula, Dicranota, Pedicia, Pilaria, *Antocha vitripennis*, Eleophila, Hexatoma
			Chironomidae	Tribe/Subfamily
			Simuliidae	Species
			Dixidae	Species
			Thaumaleidae	Species

			Ptychopteridae	Species
			Stratiomyidae	Species
			Athericidae	Species
			Rhagionidae	Species
			Tabanidae	Species

Table S5.3: In vitro single specimen tissue samples tested using the Leray primer pair prior to sequencing (Leray et al., 2013).

Species	Group
Baetidae	Ephemeroptera
Pisidium subtruncatum	Bivalvia
Caenis rivulorum	Ephemeroptera
Valvata piscinalis	Gastropoda
Athripsodes cinereus	Trichoptera
Cordulegaster boltonii	Ordonata
Bathyomphalus contortus	Gastropoda
Ecdyonurus dispar	Ephemeroptera
Luctra fusca	Plectoptera
Ecdyonurus torrentis	Ephemeroptera
Sphaerium corneum	Bivalvia
Chloroperla	Plectoptera
Hydropsyche angustipennis	Trichoptera
Rhyacophila dorsalis	Trichoptera
Limnephilidae	Trichoptera
Glossiphonia	Clitellata
Theodocus fluviatile	Gastropoda
Athripsodes aterrimus	Trichoptera
Caenis sp.	Ephemeroptera
Caenis luctuosa	Ephemeroptera
Isoperla grammatica	Plectoptera
Silo pallipes	Trichoptera
Pisidium casertanum	Bivalvia
Agapetus fuscipes	Trichoptera
Pisidium sp	Bivalvia
Margaritifera margaritifera	Bivalvia
Dikerogammarus villosus	Amphipoda
Dreissena rostriformis bugensis	Bivalvia
Dreissena polymorpha	Bivalvia
Gammarus pulex/fossarum	Amphipoda
Sericostoma personatum	Trichoptera
Sialis lutaria	Megaloptera

5.8 Supplementary information II

Tables S5.4 Venn diagram output from R, details of sOTUs in each sample type only and occurring in other sample types.

DNA – bulk only:

Baetis vernus	*Crangonyx*	*Crangonyx floridanus*
Ephemeroptera	*Gammarus tigrinus*	Glossiphoniidae
Haliplus sibiricus	*Heptagenia sulphurea*	Hydropsychidae
Hygrotus impressopunctatus	*Leptophlebia marginata*	*Limnephilus externus*
Metriocnemus eurynotus	*Orectochilus villosus*	*Orthocladius*
Oulimnius tuberculatus	*Polycentropus irroratus*	*Proasellus meridianus*
Radix balthica	*Rhithrogena*	*Rhyacophila meridionalis*
Simuliidae	*Simulium costatum*	*Simulium kiritshenkoi*
Simulium reptantoides	*Stempellinella flavidula*	Trichoptera
Trissopelopia longimana		

Microscopy only:

Acroloxus lactustric	*Agabus*	*Agapetes fuscipes*
Amphinemura	*Anabolia nervosa*	*Ancylus fluviatis*
Anisus vortex	*Aphelocheirus aestivalis*	Athericidae
Athripsodes	*Athripsodes atterimus*	Baetidae
Baetis	*Bathyomphalus contortus*	*Brachyptera*
Caenis	*Caenis luctuosa*	*Caenis rivulorum*
Ceratopogonidae	*Chaetopteryx villosa*	*Chemotopsyche lepida*
Chironomini	*Coelambus*	Corixidae
Crangonyx pseudogracilis	*Cyrnus trimaculatus*	*Dicranota*
Dreissena polymorpha	*Dreissena rostriformus bugensis*	*Dreissenidae*
Drusus annulatus	*Ecdyonurus*	*Ecdyonurus dispar*
Eleophila	Elmidae	*Elodes*
Empididae	*Ephemera*	*Ephemereliidae*
Ephemeridae	*Erpobdella testacea*	Gammaridae
Gammarus	Glossimatidae	*Glossiphonia heteroclita*
Glossiphonidae	*Glyphotaelius pellucidus*	*Goera pilosa*
Goeridae	*Gyraulus albus*	*Gyraulus crista*
Gyrinidae	*Gyrinidae larvae*	*Halesus digitatus*
Halesus radiatus	*Haliplus*	*Heptagenia*
Heptageniidae	*Hippeutis complanatus*	*Hydopscyhe pellucidula*
Hydropsyche	*Hydropsyche angustipennis*	*Hydropsychodidae*

Hydroptilidae	*Hyroptila*	*Isoperla grammatica*
Ithytrichia	*Lepdostomatidae*	*Lepidostoma hirtum*
Leptophlebiidae	*Leuctra*	*Leuctra inermis*
Leuctridae	*Limnephilus*	*Lymnaea*
Lymnaea auricularia	*Micropterna lateralis*	*Mystacides azure*
Mystacides longicornis	*Nebrioporus elegans*	*Nemoura*
Nemoura erratica	Nemouridae	*Odontocerum albicornis*
Oligochaeta	*Oreodytes sanmarkii*	*Orthocladinae*
Oulimnius	*Paraleptophlebia*	*Piscicola geometra*
Pisidium	*Pisidium casernatum*	*Pisidium subtruncatum*
Planorbis planorbis	Plecoptera	*Polycentropodidae*
Potamophylax	*Potamophylax rotundipennis*	*Prodiamesinae*
Psychodidae	Psychomyiidae	*Psychonmia pusila*
Radix peregra	Rhyacophillidae	*Rhycophila dorsalis*
Rithrogena semicolorata	*Sericostoma personatum*	*Sialis lutaria*
Silo pallipes	Simulidae	*Simulium equinum*
Sphaerium	*Sphaerium corneum*	Taenopterygidae
Tanypodinae	*Tanytarsini*	*Therodoxus fluviatilis*
Theromyzon tessulatum	Tipulidae	*Valvata cristata*
Valvata piscinalis	Valvatidae	*Veliidae*

Water – eDNA only:

Eukiefferiella	*Micropsectra notescens*	*Orthocladius fuscimanus*
Potamopyrgus oppidanus	*Unio tumidus*	

Sediment – eDNA only:

Esolus parallelepipedus	*Lumbriculus variegatus*	*Micropsectra*
Micropsectra pallidula	*Pisidium edlaueri*	

Microscopy : DNA – bulk:

Athripsodes cinereus	*Calopteryx splendens*	*Cordulegaster boltonii*
Elmis aenea	Erpobdellidae	*Haemopis sanguisuga*
Haliplus lineatocollis	*Helobdella stagnalis*	Leptoceridae
Limnephilidae	*Limnephilus lunatus*	*Pacifastacus leniusculus*
Planorbidae	*Polycentropus flavomaculatus*	*Potamophylax latipennis*

Simulium	*Simulium ornatum*	*Tinodes waeneri*

DNA – bulk : Water – eDNA:

Micropsectra atrofasciata	*Simulium reptans*

Microscopy : Water – eDNA:

Serratella ignita

DNA – bulk : Sediment – eDNA:

Ancylus	*Neureclipsis bimaculata*	*Theodoxus fluviatilis*

Sediment – eDNA : Water – eDNA:

Micropsectra contracta

Microscopy : DNA – bulk : Water – eDNA:

Baetis rhodani	*Gammarus pulex*

Microscopy : DNA – bulk : Sediment – eDNA:

Bithynia tentaculata	*Eiseniella tetraedra*	*Ephemera vulgata*
Erpobdella octoculata	*Glossiphonia complanata*	*Hydropsyche siltalai*
Limnius volckmari	*Potamopyrgus antipodarum*	Sphaeriidae

DNA – bulk : Sediment – eDNA : Water – eDNA:

Gammarus fossarum	*Prodiamesa olivacea*

Microscopy : DNA – bulk : Sediment – eDNA : Water – eDNA:

Asellus aquaticus	Chironomidae	*Ephemera danica*

Chapter 6 Discussion

6.1 Overview of main findings

The aim of this thesis was to demonstrate the applications of eDNA for freshwater macroinvertebrate monitoring. I investigated both targeted and passive approaches to detecting whole communities and high priority invasive non-native species (INNS), and also how these tools compared to current monitoring methods. In this final chapter I will summarise my main findings and discuss how it fits in to the wider context of applying molecular methods to ecological assessment.

For the first data chapter, Chapter 2, species specific primers were designed for four high priority INNS. In order to validate and test the primers, mesocosm experiments were carried out for each species, which demonstrated the sensitivity of the primers (detection within 4 hours) and how quickly the DNA degraded in the mesocosm system (between 24 hours and 1 week). Carrying out mesocosm experiments for four species (*Dreissena rostriformis bugensis*, *Dreissena polymorpha*, *Dikerogammarus villosus* and *Dikerogammarus haemobaphes*) demonstrated how using eDNA for target taxa may differ depending on the taxa being targeted. The results reinforced the need for species specific primer development to include validation steps like this one in order to demonstrate eDNA detection is effective in controlled conditions, before carrying out field studies. *Dikerogammarus* species did not produce as much DNA as Dreissenid mussels and therefore field detection may be influenced by this. For field validation, a comparison between eDNA samples and traditional kick-sampling was carried out, and for three of the four species, eDNA outperformed the established method. This demonstrated how variable current methods are for the detection of each INNS but also how efficient using eDNA can be for field detection. The fieldwork in this study was carried out in a manner which Environment Agency staff sample waterbodies for water quality, and so the results indicate a true representation of how detection of INNS could be used by monitoring staff.

In the second data chapter, the *Dreissena rostriformis bugensis* primers designed in Chapter 2 were developed for qPCR. The purpose of this chapter was also to compare the differing molecular techniques to detect *D. r. bugensis* with each other (passive vs targeted approaches) and also current kick-sample methods. Firstly, the qPCR primers were tested on the mesocosm samples generated for the first data chapter to test for a link between biomass/density and copy

number, which has been demonstrated in several previous studies. However only time since the start of the experiment was a significant factor. In field tests, both density of mussels and distance from source population had a significant relationship with both DNA copy number and metabarcoding read number. Contrary to some published studies, there was no increased sensitivity between qPCR over PCR approaches. Both targeted approaches were more sensitive than metabarcoding for detection of quagga mussels in the field, however metabarcoding still outperformed kick sampling, and had the advantage of providing information on the wider community. Therefore, these results indicate the use of molecular methods would be a valuable tool for the detection of INNS.

The third data chapter described the discovery of a previously unrecorded non-native gammarid species, *Gammarus fossarum*, in the UK by passive detection and demonstrated the processes needed in order to validate the detection of a new INNS by molecular methods. One of the important elements of this detection was that by having a good understanding of macroinvertebrate species native to the UK, I was able to query the molecular data, which may have been missed by those without taxonomic expertise. In order to verify the species identification, validation was carried out of both the reference sequences used in the database (GenBank) and of actual specimens from the rivers sampled. As the monitoring by molecular methods increases, it is hoped that such extensive verification steps may not be needed as sufficient verified reference databases will be available.

In the final data chapter, metabarcoding was applied to eDNA and bulk DNA samples collected from rivers in order to describe macroinvertebrate communities. Four sample types were analysed: kick-samples, bulk DNA samples (DNA extracted from the kick-sample) and eDNA from water and sediment within 44 sites in UK rivers. By using a metazoan specific primer pair, we aimed to detect the macroinvertebrate community, however the data showed a significant proportion of reads taken from both eDNA samples were from other sources (algae, fish, plants etc). Although the bulk DNA sample came from the kick-sample identified by microscopy a significant proportion of taxa was not picked up in the metabarcoding data, which is likely down to primer bias, biomass bias and stochasticity within the samples which were metabarcoded. Water Framework Directive classification was also run for 5 sites to demonstrate the ecological assessment information gleaned from each sample type. By using molecular methods and

traditional macroinvertebrate classification tools, a large proportion of the data gained from these new methods, such as Annelids to species level from sediment eDNA and phytoplankton assemblages from water eDNA, is in fact lost as current classification methods focus on a limited number of taxa. We therefore highlight the need to develop metrics which will utilise these novel methods, including multiple group assessment, for example the use of both algae and macroinvertebrate communities to describe water quality, which could offer a more holistic approach to ecological assessment of rivers.

6.2 Targeted or passive approaches – which road to take?

Clearly, this study shows that the choice of approach will depend on the objectives of the study or monitoring programme. If a single species is under surveillance, a metabarcoding approach would be overkill, and metabarcoding may be less sensitive than quantitative PCR approaches for species detection. So far, few studies have carried out detailed comparisons to determine which approach is most suitable. In this thesis, I demonstrated that targeted approaches are more sensitive to detection of the target species than a passive monitoring approach. However, the passive approach is able to pick up unexpected INNS, such as *Gammarus fossarum*, found as part of Chapter 4. Similar studies have also found new or unexpected species through metabarcoding of environmental or bulk DNA samples (Mahon et al., 2014; Nathan et al., 2015; Simmons et al., 2015; Brown et al., 2016). This demonstrates that although metabarcoding may be less sensitive for detection, at least when using single, widely conserved primer pairs, it has great potential for passive surveillance. If used to complement targeted approaches, metabarcoding will greatly improve ecological monitoring, including INNS detection and management. The passive approach is not as widely established as the targeted approach and therefore requires further examination of sampling strategies such as, the number of samples required to increase the sensitivity of metabarcoding for the detection of rare taxa and the use of multiple primer pairs to target different taxonomic groups.

As current approaches stand, a combination of both targeted and passive approaches would be the most powerful strategy, allowing immediate surveillance of water bodies utilising targeted primers for high priority known or anticipated species, alongside a metabarcoding approach for detection of previously unrecorded or unexpected species. This method would also generate community data, allowing impacts of INNS to be detected at community levels (Simmons et al.,

2015) and also potential ecological assessment as demonstrated by Hänfling et al., (2016). An amalgamated monitoring strategy would be of most use for INNS detection using eDNA and development of passive monitoring approaches, as summarised in Fig. 6.1.

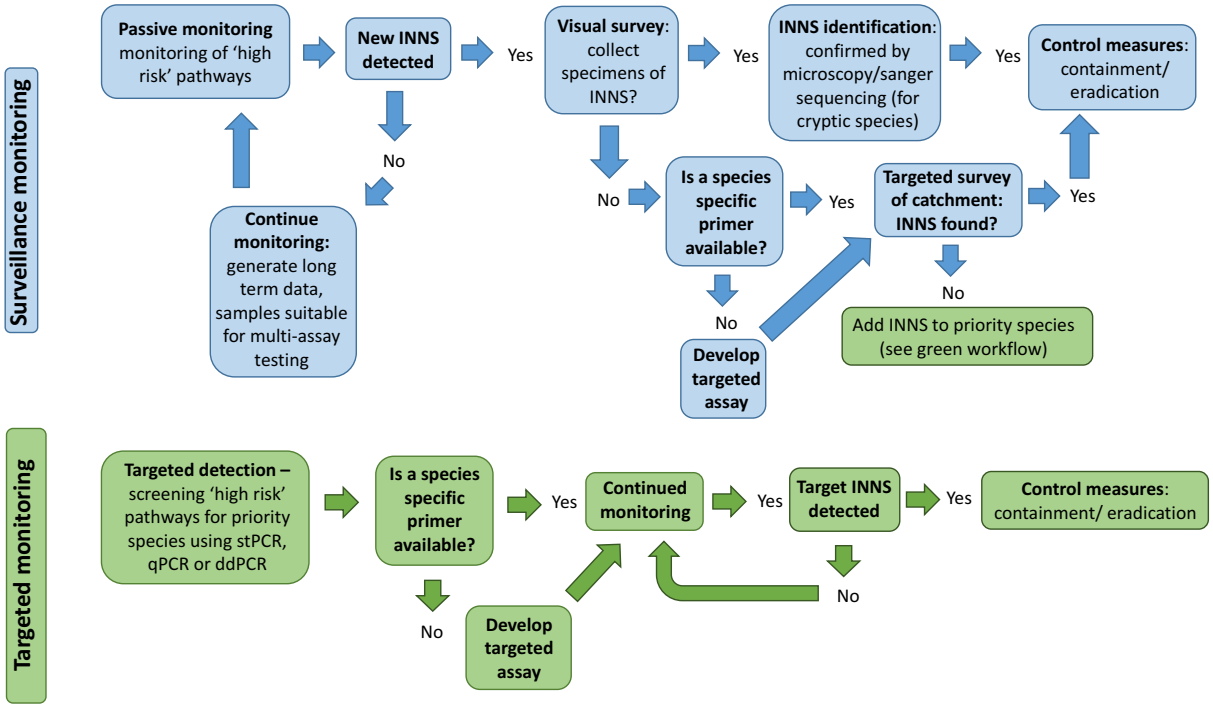


Figure 6.1: Proposed monitoring strategy using current eDNA methods. Blue: Surveillance of water bodies for INNS detection, enabling systematic screening of priority pathways and vectors (ballast water, reservoirs etc.) for INNS not highlighted by horizon scanning exercises and the collection of community data. Green: Targeted approach enabling screening of priority pathways and vectors using species specific primers and highlighting those species which require primers being developed.

6.3 What is needed from an eDNA monitoring method?

The link between research and practical application for the use of eDNA in the UK is very strong. However, what is achievable during a short-term research project and what could be implemented for a national monitoring program is quite different. During the field work for this thesis, I carried out sampling in a manner similar to standard UK Environment Agency practices, in order to make it as applicable to current sampling methods as possible. Not only did I want to make this work appropriate to end-users, I also wanted to find out what the understanding and need for eDNA monitoring methods may be to the monitoring community. I therefore carried out a small consultation of people who work with INNS but are not directly involved in academic research. The aim was to identify those who carried out monitoring work for INNS, what tools they would need to carry out their monitoring more effectively and how they considered monitoring strategies would benefit from the use of eDNA (See Supplementary Information I for full consultation questions and answers).

In total, 27 responses were received, from a variety of operational and technical staff: national government agency (16), local government (2) industry (1), research organisation (2), charity (5) and environmental consultancy (1). All participants worked on freshwater monitoring activities and the detection of target species (Q2.0 and Q3.0), including INNS and protected species, with 42.9% saying the detection of new INNS is of the highest priority within their organisation (Q5.0).

Across the participants there was an increased knowledge of eDNA and DNA methods for monitoring freshwater (51.9% limited and 48.1% good understanding) compared to marine (25.9% none, 63% limited and 11.1% good understanding) and terrestrial environments (25.9% none, 55.6% limited and 25.9% good understanding). However, on answering information with regards training needs for their organisations (Q4.3) it appears many of participants still needed, and in most cases wanted, basic training in all aspect of eDNA/DNA methods, including sampling, DNA extraction and analysis and this was reiterated with nearly half of the participants not knowing the eDNA/DNA methods being used within their organisation (Q6.1). This is not a criticism of the participant's current understandings of these methods, but rather highlights the lack of transparency within organisations to give appropriate levels of training to staff who are

directly involved in the application of these new methods. This may make them less accessible at a crucial stage where staff engagement and involvement with is fundamental to the application of these new tools in the field.

This consultation highlighted the need for training and detailed information on the current and developing methods utilising eDNA/DNA methods. A number of the participants highlighted lack of funds or resources to carry out molecular monitoring methods and stressed that in order for these methods to work in their organisations they would need to be "sold" via cost-benefit analysis. Overall, there was a clear understanding that these methods have the potential to boost INNS monitoring in freshwater ecosystems and they are worth investigating in, however I feel through this short consultation and working with colleagues in ecological monitoring most organisations would benefit strongly from training and direct communication about the research being carried out in the UK. This is key if we are to develop affordable, practical and effective eDNA/DNA monitoring schemes in freshwater.

6.4 Towards optimal protocols for ecological monitoring with eDNA

The application of eDNA for biodiversity monitoring has been rapid and its applications are likely to extend dramatically as fundamental questions such as sample method, approach (targeted vs passive) and downstream analysis are overcome. As with most new technologies, in the initial development of a tool, there has been little standardisation in terms of sample and laboratory protocols. With regards to INNS detection, the need to minimise false positives or negatives must remain a priority in future studies (Roussel et al., 2015). To ensure against false-negatives and false-positives these tools must be developed with 'gold standard' protocols in mind (Goldberg et al., 2016). Inclusion of blanks at all stages of processing (field sampling, filtering/precipitation, extraction and lab processing) and positives (such as single species tissue samples, used to establish contamination levels) has become standard in specialist eDNA laboratories. For newly detected INNS, molecular or morphological verification steps should ideally be employed, as outlined in Figure 6.1 and in Blackman et al., (2017). To allow for continued method development and robust protocols, consortiums such as DNAqua-net (Leese et al., 2016) and the US and UK DNA working groups aim to address and progress many of the issues highlighted above. This form of collaboration allows researchers from multiple disciplines to further demonstrate and advance our ability to deliver successful biodiversity monitoring using eDNA.

As discussed in the introduction and demonstrated in Chapter 3, there are several factors affecting persistence of eDNA within an aquatic environment and sensitivity levels vary depending on the molecular approach being used (Nathan et al., 2014; Simmons et al., 2015). Negative detections of a target species therefore could result from either of these factors rather than actual absence of the target species. Both biological and temporal sample replicates are needed to provide the greatest confidence in detections and avoid false results.

A largely unexplored element of eDNA availability, including by this thesis, is the effect of seasonal change. Within my thesis I focused on lotic waterbodies, where effects of changes in flow and inputs (for example: sediment point source pollution, road run off) are likely to have regular and significant short terms impacts on the DNA transport and persistence in a system. Studies have previously described effects of lake stratification and mixing on DNA availability, however, many of the impacts on river systems are still widely unknown. I carried out Spring sampling only, when flows are higher and so eDNA is likely to be more evenly dispersed throughout the water column than in drought or low flow conditions (Jane et al., 2015). However, for eDNA to be used routinely for monitoring, temporal replication is needed. Of the limited studies which have looked into seasonality effects; increased dilution of eDNA due to high flows and increased inhibition caused by increased organic matter in the water column are two examples of the possible variation which seasonality may cause (Jane et al., 2015). Similarly, it would be prudent to see if drought conditions cause eDNA to be more localised and less representative of the wider catchment. From Chapter 2 we also know individual species differ in their production of DNA, therefore the ability to detect them will vary. This variability is also likely to be increased with seasonality, as species activity and life cycle changes throughout the year, e.g. breeding season, moulting (Treguier et al., 2014; Buxton et al., 2017). To complete overall assessment of using eDNA targeted and passive approaches, further research should continue in this area to include both temporal and spatial sampling in river catchments.

Along with most eDNA studies I used PCR amplification for my samples, as the concentration of DNA from environmental samples is thought to be low. However, PCR based amplification relies heavily on the specificity of the primers being developed (both for target approaches and metabarcoding). Amplification of non-target DNA in a targeted approach or primer bias in a

metabarcoding approach will have major implications for the detection of INNS through false-positives and false negatives and has been highlighted by a number of studies (Elbrecht and Leese, 2015; Shokralla et al., 2016; Simmons et al., 2015). Therefore, a number of recent developments aim to use PCR-free detection of target species. Metagenomic or PCR-free sequencing involves the random fragmentation of all DNA in a sample. Metagenomics sequencing has been applied to invertebrate community samples (Zhou et al., 2013) and for detection of trophic interactions (Paula et al., 2014; Paula et al., 2016) but has yet to be applied to eDNA in its strictest sense. Metagenomics for eDNA could be complicated by the high proportion of non-target DNA found in the sample. However, methods such as enriching or capturing the target DNA may improve the detection rates, but so far, this technique remains challenging. For example, in an investigation of arthropod trophic interactions, enriched shotgun metagenomics detected only 15 to 239 reads out of 2 million that could be assigned to species other than the predator (Paula et al., 2016).

The Oxford Nanopore (ON) technologies MinION is also increasingly being adopted in range of studies for its high throughput and read length. Although not yet applied for targeted eDNA or biodiversity monitoring, ON technologies have already proven rapid and effective for identifying pathogen outbreaks in the field (Greninger et al., 2015; Quick et al., 2016). A similar approach could be employed for early detection of priority INNS, with the main advantages of this technology being the portability (it is suitable for use in field applications even where internet connection is limited) and the opportunity for rapid results (within 24 hours) (Votintseva et al., 2017). To test the potential of ON-MinION technology, I carried out a feasibility study in which I sequenced a selection of the bulk DNA samples generated in Chapter 5. The aim of this test was to evaluate the taxonomic resolution of Illumina MiSeq (the HTS platform used in this thesis) and the ON-MinION compared to morphological analysis. Of the 60 macroinvertebrate Families I had identified by microscopy, 50% and 46.6% were detected during from the Illumina MiSeq and MinION platforms, respectively. The sequencing methods therefore are comparable at Family level. However, MinION results were poor for higher taxonomic resolution due to the higher error rate. The MinION does however have the advantage of being able to BLAST sequences in real time. During the first hour of the run I identified several common taxa, including the new non-native *Gammarus fossarum*, which demonstrates the potential this technology has for field application compared to the Illumina platforms.

High Throughput Sequencing (HTS) is heavily reliant on bioinformatics and reference databases. There are a number of automated methods which can be adapted for a wide range of datasets (Caporaso et al., 2010 – Qiime; Boyer et al., 2016 – Obitools) and also those which have been designed for a specific set of data (Hänfling et al., 2017 – metaBEAT). In this thesis two different bioinformatics pipelines were employed. Firstly, the PIPCOI pipeline (Gweon et al., unpublished data) in Chapter 2, and secondly in Chapters 4 and 5, metaBEAT (Hänfling et al., 2017) with minor modifications for the macroinvertebrate dataset. Overall, these pipelines carry out the same broad steps: de-multiplexing and pairing reads, chimera detection and removal, clustering and taxonomic identification of reads (Boyer et al., 2016). However, within pipelines, algorithms for different steps are sometimes different (for example: sequence clustering using VSEARCH (Rognes et al., 2016) or DEBLUR (Amir et al., 2017) and comparisons between bioinformatics approaches are not common practice. In this study, we used BLASTN which is a similarity-based approach to assign taxonomic identification. However, Coissac et al., (2012) highlight four alternative approaches in their review, indicating the variation of approaches within each step. Comparisons of bioinformatics approaches should be carried out in order to determine differences in data outputs from these approaches and automated pipelines. Similarly, as described by Coissac et al., (2012) the main objective for biologists using HTS data is to assign unknown DNA sequences to a reference database, with the data output being heavily reliant on the accuracy of the sequences. During my thesis, I discovered a number of errors within publically available reference databases and therefore fully endorse the verification of reference sequences before they are published.

6.5 Conclusion

Significant progress has been made in this field since the first use of eDNA to detect an INNS less than 10 years ago (Ficetola et al., 2008), including the development of a large number of species specific assays and progress in community-based analysis. This thesis shows that the development of species specific primers may require a great deal of investment in terms of primer design, optimisation and validation, but the need for these tools is prevalent, and in the case of high priority INNS such as those identified via 'horizon scanning', targeted eDNA approaches may be the most suitable method of early detection. HTS approaches also provide unprecedented opportunities for passive surveillance and understanding community structure, but further work

is needed to improve their sensitivity and maximize the amount of information we can obtain from this powerful technology.

Some fundamental elements of eDNA monitoring methods still remain largely unresolved and widely unregulated in terms of sampling strategy or analysis. This is due not to a lack of research but rather the huge variation in both systems and species being studied. There is a strong need to develop appropriate validation processes and consistent field and laboratory standards, but without stifling progress as the techniques are progressing so rapidly. In the immediate future, I think efforts should focus on the effects of seasonality of eDNA detection of species, as this is of vital importance and will determine the application of eDNA for routine monitoring use. I also feel the use of a multi-primer HTS assays to test eDNA samples taken from rivers is likely to allow a much greater biodiversity assessment, and if we were able to assess numerous taxa together this will offer a greater certainty and level of biodiversity assessment to regulators and researchers alike. Perhaps most importantly, if we are to use eDNA for monitoring communities or the detection of new INNS, targeted or community based, we need to have an understanding of both the molecular methods and the species which we are trying to detect. Whether we target a fish species in a lake or a mussel in a river, to use this method effectively, ecological understanding must play a key role in developing new tools.

6.6 Supplementary Information I

Table S6.1 Questions and answers from the eDNA INNS consultation carried out in March 2017

1.0 Name
1.1 Email address
1.2 Phone number
1.3 Job title
1.4 Brief job description
1.5 What organisation do you work/volunteer for
National government agency 59.2% Local government 7.4% Industry 3.7% Research organisation 7.4% Charity 18.5% Environmental consultancy 3.7%
2.0 Which of the following environments fall within the remit of YOUR ORGANISATION? Please tick as appropriate and include all that apply.
Freshwater (surface) 100% Groundwater 66.7% Marine 77.8% Terrestrial (surface) 85.2% Soil 59.3% Air 48.1% Other: Soil/Air permits 3.7%, Geology/Caves 3.7%, Pathogens 3.7%
2.1 Which of the following environments fall within the remit of YOUR ROLE? Please tick as appropriate and include all that apply.
Freshwater (surface) 92.6% Groundwater 7.4% Marine 33.3% Terrestrial (surface) 55.6% Soil 11.1% Air 11.1% Other: Pathogens 3.7%
3.0 Which of the following approaches are used in the monitoring activities carried out by your organisation? Please tick as appropriate and include all that apply.
Animal community diversity estimates 74.1% Option Plant community diversity estimates 63% Microbial community diversity estimates 29.6% Detection of individual target species (i.e. invasive species, conservation priority species) 100% Detection of specific eukaryotic parasites and parasitoids 40.7%

Detection of specific bacterial and viral pathogens 44.4% Detection of not previously recorded species (invasives, parasites, pathogens) 74.1% Others: Terrestrial mapping 3.7%, Air quality monitoring 3.7%, Marine protected area 3.7%
3.1 What are the reasons for carrying out this monitoring?
International directive 29.6% European directive 77.8% National directive 70.4% Regulation 55.6% Local investigation/monitoring needs 81.5% Consultation 33.3% Local conservation issues 66.7% Local recording/monitoring 59.3% Other:
4.0 What is your level of understanding of the application of DNA/eDNA to monitor biodiversity in freshwater environments?
Limited 48.1% Good understanding 51.9%
4.1 What is your level of understanding of the application of DNA/eDNA to monitor biodiversity in marine environments?
None 25.6% Limited 63% Good understanding 11.1%
4.2 What is your level of understanding of the application of DNA/eDNA to monitor biodiversity in terrestrial environments?
None 25.6% Limited 55.6% Good understanding 18.5%
4.1 Further to your answer to question 4.0, 4.1 and 4.2 please give details, including identifying any training needed for yourself or organisation (including sampling method, lab work and data analysis)
Only done eDNA for GCN - companies standard approach is to outsource everything to consultants and the specification for the works usually has to align with standard EA measures - e.g, APEM may be contracted to do macrophyte and invertebrate surveys, and the survey spec has to follow UKTAG guidance for RIVPACS/LEAFPACs Broader knowledge required to understand the science underpinning DNA / eDNA monitoring, so that its application and usefulness and limitations are better understood. More detailed knowledge required around sampling collection and how to collect good samples. Less knowledge required about the lab process, apart from a few individuals that will manage the contracts for this analysis and so need in-depth understanding of the analysis methods to enable contract management.

Specific knowledge relating to bioinformatics required to make the most from these data.
Simple techniques to be used by staff
DNA in wetland habitats for the potential presences of American bullfrog
Nothing specific because I work in collaboration with specialised colleagues
Undertaken eDNA on great crested newts when at a consultancy. Last year I assisted a PhD student from Swansea uni for sampling of signal crayfish and Chinese mitten crabs within the Dee to help confirm distribution
Used genetic techniques for non-invasive mammal monitoring in Interreg funded project 'Mammals in a Sustainable Environment'. Included material collected from hair tubes, scats and bait pots (to collect droppings)
Need more information about how it can be used and its limitations. Has to be cheap, fundable and easy to use by staff and volunteers. Otherwise won't be used.
eDNA would save an enormous amount of time and help with future control work.
Method suitability, method limitations, sampling method, lab work and data analysis
I / we have heard of eDNA and some of it's applications but this new and only basic knowledge. Training would be required in all aspects
I am aware of eDNA as a monitoring tool, and have participated in some eDNA sampling for great crested newts, but that is the extent of my knowledge. I would find it extremely useful to receive training on the application of eDNA as a monitoring tool for all environments, and in sampling techniques and data analysis. In my role, I would not need to know the specifics of the laboratory analysis. I would think the above applies to a great many of my colleagues involved in the management of freshwater, marine and terrestrial environments.
NRW plan to trial eDNA for diatoms in 2017 (alongside microscopy technique), and FERA giving NRW a talk 29th March 2017.
Development of standard sampling methods and associated training, specifying laboratory techniques for contracts, interpreting data and applying to make conservation decisions.
further NGS data analysis training required
sampling methodology
Natural England do not have labs/ability to analyse data so this would be contracted, however would be useful to understand the methodology and limitations to help inform any future monitoring opportunities.
sampling method for area staff, lab work if EA labs take on analysis. Data interpretation and how varies from std monitoring methods
I have gleaned some knowledge of this topic through attending workshops and reading internal documents. As far as I'm aware NRW has not undertaken monitoring of this type other than collecting a few samples for the EA diatom DNA trial.
CRT would need to have the benefits of DNA/eDNA "sold" and the benefits over the existing methods of monitoring currently in place for our requirments.

I am NE's marine lead for eDNA. We are quite new to the area but catching up quickly. I think being able to participate in workshops and meetings with partners will be cruicial for staying up to date with developments in this fast advancing field.
Sampling and field collection protocols; use of sampling equipment; how to preserve samples for DNA extraction; data analysis and interpretation; Bioinformatics
5.0 How important is the detection of new invasive alien species within your organisation? (0 not relevant, 4 highest priority)
0 – 0% 1 – 0% 2 – 18.5% 3 – 40.7% 4 – 40.7%
5.1 How important is the continued monitoring of invasive alien species within your organisation (0 not relevant, 4 highest priority)
0 – 0% 1 – 3.7% 2 – 18.5% 3 – 40.7% 4 – 37%
5.2 Please provide a brief description of any invasive alien species monitoring activities carried out by your organisation (species, approximate number of sites and samples per year). Indicate whether relative or absolute abundance is assessed.
Zebra mussel plates at reservoirs, once every 5 years major surveys across c.150 sites (fish/aquatic invert/botanical), monitoring done at around 100 locations alongside standard 3 min kicks as drought plan baseline monitoring data. Fund riparian monitoring across catchments through stakeholder groups
Most of our monitoring programmes incidentally monitoring IAS. We have targeted surveys for Dikerogammarus villosus and species that we are engaged in eradicating, such as Ludwigia grandiflora and topmouth gudgeon
Invasive shrimp monitoring - circa 200 sites per anum. Other INNS are recorded as part of wider monitoring activities but not specifically targeted.
See alert system delivered through NNSS
We have a small direct monitoring programme for Priority INNS as well as a much larger indirect programme (where the detection of INNS is not targeted but they are recorded as part of a wider survey) - we also work closely with external organisations to share our records with others.
Japanese knotweed-updating Gower AONB survey, previous county wide surveys in 1992 and 1998. Planning applications Him' balsam-on going mapping G hogweed-all known sites mapped Himalayan knotweed-sites mapped Cordgrass (spartina)-annual survey Hottentot fig-annual survey fauna eg terrapin, parakeet-records updated
Multiple species and many locations nationally. Most common freshwater and terrestrial invasive species.

Relative and absolute abundance of invasive fish species in several Cumbrian lakes

For plants we and volunteers monitor and record INNS throughout summer, so we can manage them accordingly. Animals are a lot more difficult, and we tend to rely upon volunteers from the fishing and canoeing communities who notice them. Mink we monitor as part of North Wales otter survey.

For mammals: surveillance to identify presence of grey squirrel and mink at sites across Wales to implement control. Likewise, for deer INNS

Relative abundance across the 520 square miles of park in cooperation with other partners. Data kept by LRC much of which is opportunistic data--reports from public. We track data and do what we can where we can but ultimately responsibility of landowner. No exact numbers available. More fundamental needs than this for us. some active monitoring of Japanese knotweed and other aquatic invasive but limited in scope and geography.

40km's of river bank surveyed annually. Watercourse surveys for invasive inverts, musseks and crayfish carried out

Crayfish plague (via CEFAS), site condition assessments for notified sites

5 Invasive non-native plant species are currently controlled and monitored as evidence for their control and our methods. Surveying of invasive plants is done over the whole Medway catchment which equates to about 200km of riparian habitat.

No or very limited INNs specific monitoring, but INNS should be picked up as a negative indicator (cause for concern) in condition monitoring of protected sites. Or specific surveys commissioned for particular sites We are reliant on others monitoring eg Environment Agency

Many species are monitored, either via targeted monitoring, as incidental records of routine monitoring or as part of specific site management e.g. NNRs or SSSIs. Some of this work is carried out via partner organisations e.g. the Wildlife Trust and BASC help with mink and grey squirrel control. I cannot provide details of the number of sites and samples etc, but hopefully you will have responses from colleagues who will provide this information. Species that we will have monitoring data for include (this is not an exhaustive list): Japanese Knotweed Himalayan Balsam Giant Hogweed Rhododendron Sea Buckthorn Crassula Parrot's Feather Killer Shrimp-we do carry out specific targeted monitoring for this species-see response from my colleague Mel Lacan. Signal Crayfish Mink Top Mouth Gudgeon Grey squirrel (in our red squirrel project areas).

Targeted invasive shrimp monitoring carried out annually (approx. 30 sites per year in S Wales), giving absolute abundance. Although 2017 resources targeted on biosecurity awareness raising.

Limited

using DNA/environmental DNA currently only Didemnum vexillum, bi-monthy, currently 6 sites, new project starting in April 2017 looking at NGS methods to target all relevant marine INNS, targeting major hotspots for INNS introduction. One of the main focusses will be ascidians, others will be included in the duration of the project.

Topmouth Gudgeon, 6 sites once a year Killer shrimp 4 sites a year

Targeted marine INNS surveys do not take place on a national level but there may be local level surveys linked to assessing the condition of sites. In general, any

INNS found on any surveys are recorded and highlighted on marine recorder as this is written into all contracts.

Dikerogammarus villosus ~70 sites last time I checked. Some D. Haemobaphes. Rangia - localised. All are just qualitative presence absence surveys. Also detection of invasive in routine monitoring (KS, Fish, plant etc).

1. After the first discovery of D. villosus in 2010, monitoring at potential sites in Wales identified two sites which had been heavily colonised - Cardiff Bay and Eglwys Nunydd Reservoir. This was followed by intensive monitoring and investigation to check for pathways of spread. Now that it is clear that the animals cannot colonise further unaided, the focus is in biosecurity and checking potential new sites. The amount of effort has certainly declined since the initial discoveries. 2. There has been limited monitoring for D. haemobaphes but it appears not to have reached Wales yet. 3. Ponds where Ludwigia has been treated, are visited regularly to check from re-growth. 4. During routine freshwater samplin,g records are made of any invasive species seen such as Himalayan Balsam and Japanese Knotweed. Similarly any invasive species found incidentally in samples will be recorded.

Japanese Knotweed, Himalayan Balsam, Giant Hogweed, Floating Pennywort, Parrots Feather, Water Fern, Duckweed, Common Ragwort, Signal Crayfish, North American Mink, Zebra Mussel.

As a part of our marine monitoring framework, we require all our contractors to note down all the invasive sp they come across durign surveys and flag them up to us. Our INNS working group promotes working with partners to tackle marine invasive species, but at the same time we recognise there is very little we can do in terms of management if and when they are found on designated sites. Our marine condition assessment process is looking into this as a critical issue.

Currently we do not have a dedicated monitoring or surveillance programme for IAS in Natural England. As part of routine monitoring of protected sites (e.g. Common Standards Monitoring) we will record negative indicators on site condition such as the presence of high impact IAS. We have ad hoc monitoring at some sites with known IAS problems e.g. floating pennywort on Pevensey Levels (mapping distribution/relative abundance etc). In 2011/12 we carried out monitoring of 16 freshwater SSSIs for the presence of Dikerogammarus villosus when the species was first found in the UK (Graftham Water).

5.3 Does your organisation have a list of target invasive alien species which are already present in the UK? If yes, please list.

Yes (2)

Zebra Mussel, Water primrose, Asian clam, floating pennywort, Japanese knotweed, quagga mussel, giant hogweed, giant knotweed, Himalayan balsam, water fern, killer shrimp, demon shrimp (list changes as part of a risk matrix tool we have that updates relative risk depending on findings of surveys and reports of findings from other areas)

yes, too extensive to list

Yes, I don't have access to this. Contact Alice Hiley.

yes

I use the Non-native Species Secretariat list and watch other lists eg Water Directives and EU lists

Japanese knotweed, giant hogweed, floating pennywort, azolla the main plant species. We have not yet developed the list for animals, but is likely to include the invasive mussel species, the invasive shrimp species, non-native crayfish species.
It does but too large to list here
Yes part of NNSS local action group, so am well informed in this area, obviously more aware of INNS within catchment, but mechanisms in place for Horizon scanning as to what may be making its way into our catchment.
For mammals only - grey squirrel, mink, muntjac, sika deer and Chinese water deer.
yes. See list for Wales.
Zebra mussel, chinese mitten crab, himalayan balsam, japanese knotweed , giant hogweed, floating pennywort , Asian hornet
Yes, too numerous to list here. Work with GBNNSS. Also considers locally non-native
Yes, via NNSS
For freshwater, the WFD high impact list and GB non-native species alert system are the lists used primarily
Not sure if I understand this question. We tend to use GB-wide resources such as NBN gateway. There is also a Wales Priority list, some of which are already present, some which are not. We do have local knowledge of what species we have and where they are.
probably - I use the GBNNSS website
I don't think so
Scottish National Heritage and GB NNS secretariat hold a current list of INNS. The priority species for our organization are the species with currently only limited spread such as Didemnum vexillum, and other ascidians, crustaceans
Killer Shrimp, Zebra mussel, Topmouth gudgeon and signal crayfish
We use the MSFD list developed by CEFAS for Defra
I'm not sure what this means. We work to the INNS Directive listing
Japanese Knotweed, Himalayan Balsam, Giant Hogweed, Floating Pennywort, Parrots Feather, Water Fern, Duckweed, Common Ragwort, Signal Crayfish, North American Mink, Zebra Mussel.
yes, but sorry I don’t have the list. Our INNS lead is Jan Maclennan.
Yes - risk assessment produced by GB NNSS and WFD impact list. We have identified eDNA monitoring/research priorities for freshwater IAS with groupings around 1. amphibians/ reptiles; 2. Mitten crab; 3. Crayfish spp.; 4. Ponto-Caspian Alert species; 5. Molluscs; 6. Aquatic plants; 7. Fish
5.4 Does your organisation have a list of priority invasive alien species which are not present in the UK? If yes, please list
No (2)
Yes (2)
We have a multimetric system for prioritising IAS, which inserts new species at an appropriate ranking within the priority order
Yes, I don't have access to this. Contact Alice Hiley.
Yes - see Roy et al 2014 GCB

Not yet
It does but too large to list here
Yes, again see above
Mammals - Herpestes javanicus, Small Asian mongoose Myocastor coypus , Coypu Nasua nasua, South American coati Nyctereutes procyonoides, Raccoon dog Procyon lotor, Raccoon Sciurus niger, Fox squirrel Tamias sibiricus, Siberian chipmunk
yes. See Wales priority list.
Many
Yes, work with GBNNSS
No but I imagine I could get this from NNSS
Use WFD alarm list and GB non-native species secretariat/ NNSIP horizon scanning information
Yes. We use the List of Union Concern and the GB Top 30 Horizon scanning list to identify the high-risk species which are not yet present in the UK.
don't know
I don't think so
same as above, the main species from the list will be mitten crab, slipper limpet
Gyrodactylus salaris
Again we use the MSFD list developed by CEFAS for Defra
Not that I'm aware of.
yes, but sorry I don't have the list. Our INNS lead is Jan Maclennan.
Yes - UK and EU Horizon scanning for IAS
5.5 Does your organisation have a list of invasive alien species which are present in the UK which it no longer monitors? If yes, please list
No (6)
no
Yes, I don't have access to this. Contact Alice Hiley.
While the priority species list for direct monitoring can change, all species found in a given survey are recorded.
Himalayan balsam
Yes, although always looking into methods to assist in monitoring eg Edna testing for mitten crabs and signal crayfish, which we didn't have a robust methodology to monitor. Currently looking into methods for quagga & zebra mussels that are not currently known to be in catchment, but I am sure will be soon.
no.
Many not monitored directly or on ad-hoc basis
No, as we do not specifically monitor all INNS
No. I expect there are plenty of things that we don't formally monitor, as it's not possible to monitor for everything, but we don't have a list. Examples may be things like grey squirrel, buddleia, Japanese knotweed, Himalayan balsam. They are so widespread that we don't carry out a full monitoring regime for them, they are just noted during routine surveys.

don't know
N/A
Not a formal list but there are species we know are here which we don't have targeted monitoring for, eg Signal crayfish, plants many others. Will record if find in routine monitoring though.
Not that I'm aware of.
Himalayan balsam, ragwort
yes, but sorry I don't have the list. Our INNS lead is Jan Maclennan.
NA
5.6 If you answered yes to 5.5, please select reasons why
Eradication not possible 54.5%
Significant established populations 81.8%
Cost 90.9%
Suitable method 9.1%
Political will 9.1%
Other 9.1%
5.7 If eDNA were used for invasive species detection, what would be the priority monitoring for your organisation (please circle 3 only)
Early warning 92.3%
Continued monitoring of spread 65.4%
Successful eradication 57.7%
Pathway monitoring 38.5%
Kit checking/screening (Inc boats) 11.5%
Pathogen/Virus detection 11.5%
Cost effective tool 38.5%
A non-invasive method of sampling 30.8%
Other: Unsure – still under consideration 3.8%
6.0 Does your organisation currently utilise DNA based monitoring approaches? (This refers to both DNA of an organism and DNA collected from environmental samples (eDNA))
Yes 63%
No 37%
6.1 If you answered "yes" to 6.0, which techniques do you personally know that your organisation uses. Tick as appropriate from the list below
PCR for species detection 25%
qPCR for species detection 25%
DNA barcoding 43.8%
Population genetic approaches (microsatellites, SNP's etc.) 18.8%
eDNA, metabarcoding (NGS based community barcoding) 31.3%
metagenomics 6.3%
NGS – Illumina 12.5%

NGS – PacBio 0%
NGS – IonTorrent 0%
NGS - Oxford Nanopore 6.3%
T-RFLP 6.3%
I don't know 43.8%
Other 12.6%
6.2 Please provide details regarding the specific monitoring needs these methods are applied to.
Don't know the specific method - eDNA I believe for GCN but method as specified in Natural England guidance
we are in the process of incorporating Edna techniques into our monitoring programme
using metabarcoding to identify diatom species present in a sample - direct replacement for microscopy analysis as part of wfd classification process. Have to be able to identify diatom taxa and provide a figure of relative abundance for diatom taxa to enable calculation of biotic indices
Freshwater Diatoms
Likely to be for GCNs
I cannot answer this fully for CEH, but my own interests focus on freshwater fish
Chinese mitten crabs and signal crayfish
These were used as part of the MISE project (noted above) and were undertaken at the Waterford Institute of Technology. Methods were used to: otters - genotype spraint samples to estimate population size and distribution; squirrels - hair tube to identify species presence and red squirrel population size/haplotype; harvest mouse - novel technique using bait pots to collect droppings to identify species presence.
Don't understand enough about methods to comment
Great Crested Newt presence/absence, information used in relation to licence applications. fish community information in SSSI lakes Vegetation community information - comparing traditional quadrat method with dna based method
Surveillance of great crested newts, monitoring soil microbial diversity, developing applications for targeted species detection and characterisation of communities and functional groups.
same as 5.7 above?
Freshwater diatom
as above
protected sp detection, examining more cost-effective monitoring methods for sp assemblages,
6.3 Are there any applications of DNA based methods you would like to see implemented to assist current monitoring programmes? If yes, please state
Would be keen to see anything for invasive species, particularly invertebrates
better use of eDNA for IAS screening
Yes - for detection of invasive species arrival and spread
Yes. I would suggest that the use of DNA-methods for INNS offers a huge amount of scope - especially for early detection. There are also applications for DNA-

methods in places where conventional monitoring is difficult - such as in the survey of lake fish.
Alien fish, mussel, crayfish, shrimp population survey/spread
Metabarcoding for freshwater fish
Yes mussels
Would like to continue to use these techniques but don't have in-house capacity.
Yes, aquatic invert sampling
Many. Rare fish presence, fish community assessment, fish biomass assessment, crayfish plague spores, native and non-native crayfish
I would like to investigate using eDNA to ascertain how far up the catchment Mitten Crabs spread.
Top Mouth Gudgeon is currently a high priority for us here in Wales. Also, it may be very useful as an early warning tool for some species which we don't yet have in Wales e.g. Quagga Mussel, Spiny Cheeked Crayfish.
spring 2017 to start eDNA diatom monitoring
Application for characterising communities and ecosystems in good condition or ecological health.
environmental DNA allowing for population genetic studies and network analysis
invasive species detection
Zebra mussel (Dreissena polymorpha), Quagga mussel (Dreissena rostriformis bugensis), Killer shrimp (Dikerogammarus villosus) and Demon shrimp (Dikerogammarus haemobaphes).
no
yes- subtidal sediment monitoring, INNS monitoring, lamprey monitoring (especially with regards to access to spawning sites up river)
6.4 Are there any monitoring challenges you are currently unable to meet because the current tools are not fit for purpose which you feel DNA based methods may allow you to achieve?
We manage water across Yorkshire on a grid system - water can theoretically be taken from any river in the county and sent to any location which obviously poses significant pathways for potential INNS transfer - quick methods to help assess risk would be useful (e.g, water taken from river may go into a settling lagoon for a day before being pumped - that would theoretically give a day in which should a sample show an INNS is present then the pumping could be stopped)
IAS screening
Lake fish population assessment. Lake phytoplankton analysis has real data quality issues doe to a lack of UK based analytical skills
Yes - as above, the application of DNA-based methods to supplement/replace conventional INNS techniques offers the potential to make our current programme much more effective by covering a much larger spatial scale. There are also situations where we have difficulty undertaking effective monitoring due to environmental/other constraints and DNA based methods would be of great use - either to supplement or replace conventional methods
Spatially extensive monitoring is currently not possible due to limited resources and other limitations, although current methods are fit for purpose

Yes, this was CMC for us, we have previously had funding for methods of monitoring and capturing them, which failed on all accounts, hence Edna our current last hope. Methods used elsewhere did not work on the Dee due to numerous designations and that the traps had to be non-target species friendly - and nothing else worked. Similarly, the mussels, as landowner permission for monitoring plates etc proving difficult to obtain.
Cost-effective and rapid techniques to identify presence of grey squirrels, particularly in areas where they have been (or are near to being) eradicated and where red squirrels are also present.
All aquatic sampling
Many. Rare fish presence, fish community assessment, fish biomass assessment, crayfish plague spores, native and non-native crayfish
See 6.3
Cost effective and consistent means of sampling for INNS species that could be used by non-experts. ie not relying on expert(s) for identification
Top Mouth Gudgeon eradication programme-e DNA would be extremely useful to help us ascertain if eradication has been successful. I am aware that the use of eDNA for this species is already available, and we have contacted CEFAS to explore the possibility of using e DNA to assist our TMG programme. Also, I suspect it would be a useful tool in Dikerogammerus monitoring-more cost effective than the time-consuming kick sampling and trapping techniques?
eDNA for invasive shrimp - if cost effective would be helpful to get better coverage and presumably more accurate coverage
sensitivity of metabarcoding community based approaches to allow early detection, issues with contaminations, false negatives, still high cost, need further training in data analysis, issues with quantifications,
Lake fish surveys. Deep water fish and invertebrate sites, where boat access difficult
The above plus possibly signal crayfish.
We would need specialist advice on what methods were available and that would benefit CRT and detail of actual costs for potential services provided.
lamprey example above. Understanding subtidal sabellaria better? Understanding how infaunal sediment communities respond to change (biotope approach doesn't work very well)
Current monitoring and sampling techniques for freshwater species (fish, inverts, IAS).
7.0 Please add any other comments you would like to make
Our current position is we are keen to utilise DNA based monitoring methods, but would only do so should they be accepted by our regulators (i.e EA), so keen to see formal position statement issued such as Natural England did in regard to e DNA for herptofauna
Rosie for president.....
I mainly deal with large terrestrial plants so don't really know if the DNA applications are suitable but interested as I also undertake general INNS/ IAS training courses
Apologies for brevity of some of the above answers and for my incomplete knowledge of total CEH interests in Edna methods

NB - my answers purely relate to mammalian related work within NRW.

Good luck with your studies. Please do contact me if NRW can be of any further assistance. Hilary

Need for sampling invasive species in still water, particularly those that we struggle to detect using conventional methods like topmouth gudgeon

As an organisation, we are attempting to keep up with developments and to identify monitoring needs that could could be fulfilled using molecular methods. We are in closest contact with Simon Creer and his team at Bangor University.

Our work is contracted out to NatureMetrics

See attached work tab for examples of the range of eDNA work NE is involved with. Need for open source eDNA based methods - commercialisation by some organisations is certainly a limiting factor on adoption of new approaches to monitoring. Confidence needed in reference databases e.g. gen-bank

References

Alberdi A., Aizpurua, O., Gilbert, T.P and Bohmann, K. (2018) "Scrutinizing Key Steps for Reliable Metabarcoding of Environmental Samples." Methods in Ecology and Evolution / British Ecological Society 9 (1):134–47.

Aldridge, D.C., Elliott, P. and Moggridge. G.D. (2004) "The Recent and Rapid Spread of the Zebra Mussel (*Dreissena Polymorpha*) in Great Britain." Biological Conservation 119 (2):253–61.

Aldridge D.C. (2013) GB non-native organism rapid risk assessment for *Dikerogammarus haemobaphes* (Eichwald, 1841).

Aldridge, D. C., Ho, S. and Froufe. E. (2014) "The Ponto-Caspian Quagga Mussel, *Dreissena Rostriformis Bugensis* (Andrusov, 1897), Invades Great Britain." Aquatic Invasions / European Research Network on Aquatic Invasive Species.

Amir, A., McDonald, D., Navas-Molina, J.A., Kopylova, E., Morton, J.T., Xu, Z.Z., Kightley, E.P. et al. (2017) "Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns." mSystems 2 (2).

Anderson, L G., Dunn, AM., Rosewarne, P.J. and Stebbing. P.D. (2015) "Invaders in Hot Water: A Simple Decontamination Method to Prevent the Accidental Spread of Aquatic Invasive Non-Native Species." Biological Invasions 17 (8):2287–97.

Ardura, A, Zaiko, A., Martinez, J.L., Samuiloviene, A., Semenova, A and Garcia-Vazquez. E. (2015a) "eDNA and Specific Primers for Early Detection of Invasive Species--A Case Study on the Bivalve Rangia Cuneata, Currently Spreading in Europe." Marine Environmental Research 112 (Pt B):48–55.

Ardura, A, Zaiko, A., Martinez, J.L., Samuiloviene, A., Borrell, Y. and Garcia-Vazquez. E. (2015b) "Environmental DNA Evidence of Transfer of North Sea Molluscs across Tropical Waters through Ballast Water." The Journal of Molluscan Studies 81 (4). Oxford University Press:495–501.

Bacela-Spychalska, K, and Van Der Velde. G. (2013) "There Is More than One 'killer Shrimp': Trophic Positions and Predatory Abilities of Invasive Amphipods of Ponto-Caspian Origin." Freshwater Biology 58 (4). Blackwell Publishing Ltd:730–41.

Bai, Y, Liang, J., Liu, R., Hu, C and Qu. J. (2014) "Metagenomic Analysis Reveals Microbial Diversity and Function in the Rhizosphere Soil of a Constructed Wetland." Environmental Technology 35 (17-20):2521–27.

Barnes, M A., and Turner. (2015) "The Ecology of Environmental DNA and Implications for Conservation Genetics." Conservation Genetics 17 (1). Springer Netherlands:1–17.

Barnes, M.A., Turner, C.R., Jerde, C.L Renshaw, M.A., W. Chadderton, L. and David M. Lodge. (2014) "Environmental Conditions Influence eDNA Persistence in Aquatic Systems." Environmental Science & Technology 48 (3). American Chemical Society:1819–27.

Bij de Vaate, A., K. Jazdzewski, H. A. M. Ketelaars, S. Gollasch, and G. Van der Velde. (2002) "Geographical Patterns in Range Extension of Ponto-Caspian Macroinvertebrate Species in Europe." Canadian Journal of Fisheries and Aquatic Sciences. Journal Canadien Des Sciences Halieutiques et Aquatiques 59 (7):1159–74.

Bista, I, Carvalho, G.R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., Christmas, M. and Creer. S. (2017) "Annual Time-Series Analysis of Aqueous eDNA Reveals Ecologically Relevant Dynamics of Lake Ecosystem Biodiversity." Nature Communications 8 (January):14087.

Blackman, R.C., Constable, D., Hahn, C., Sheard, A., Durkota, J., Hänfling, B. and 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 (2):177–89.

Blackman, R.C, Hänfling, B, and Lawson Handley. L. (2017b) The use of environmental DNA as an early warning tool in the detection of new freshwater invasive non-native species. *In review*

Blackman, R.C, Hänfling, B, and Lawson Handley. L. (2017c) Targeting the invader - species specific detection of priority non-native species using environmental DNA in UK rivers. *In review*

Bohmann, K, Evans, A., Gilbert, M, Carvalho, G.R., Creer, S., Knapp, M., Yu, D.W. and de Bruyn. M. (2014) "Environmental DNA for Wildlife Biology and Biodiversity Monitoring." Trends in Ecology & Evolution 29 (6):358–67.

Bolger, A.M., Lohse, and Bjoern Usadel. (2014) "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." Bioinformatics 30 (15):2114–20.

Bonada, N.N.P, Resh, V.H and Statzner. B (2006) "Developments in Aquatic Insect Biomonitoring: A Comparative Analysis of Recent Approaches." Annual Review of Entomology 51:495–523.

Borrell, Y.J., Miralles, L., Do Huu, H., Mohammed-Geba, K. and Garcia-Vazquez. E. (2017) "DNA in a bottle—Rapid Metabarcoding Survey for Early Alerts of Invasive Species in Ports." PloS One 12 (9). Public Library of Science: e0183347.

Bovy, H.C., Barrios-O'Neill, D., Emmerson, M.C., Aldridge, D.C and Dick. J.T.A (2015) "Predicting the Predatory Impacts of the 'demon Shrimp' *Dikerogammarus Haemobaphes*, on

Native and Previously Introduced Species." Biological Invasions 17 (2). Springer International Publishing:597–607.

Boyer, F, Mercier, C., Bonin, A., Bras, Y.L., Taberlet, P. and Coissac. E. (2016) "Obitools: A Unix-Inspired Software Package for DNA Metabarcoding." Molecular Ecology Resources 16 (1):176–82.

Brolaski, M.N., Venugopal, R.J. and Stolow, D. (2008) Kits and processes for removing contaminants from nucleic acids in environmental and biological samples. USPTO 7459548.

Bronnenhuber, J.E., and Wilson. C.C (2013) "Combining Species-Specific COI Primers with Environmental DNA Analysis for Targeted Detection of Rare Freshwater Species." Conservation Genetics Resources 5:971–75.

Brown, E.A., Chain, F.C.C., Zhan, A., MacIsaac, H.J., and Cristescu. M.C. (2016) "Early Detection of Aquatic Invaders Using Metabarcoding Reveals a High Number of Non-Indigenous Species in Canadian Ports." Diversity & Distributions, July. https://doi.org/10.1111/ddi.12465.

Buxton, A.S., Groombridge, J.J., Zakaria, N.B. and Griffiths R.A. (2017) "Seasonal Variation in Environmental DNA in Relation to Population Size and Environmental Factors." Scientific Reports 7 (April):46294.

Cai, W, Ma, Z., Yang, C., Wang, L., Wang, W., Zhao, G., Geng, Y. and Yu. D (2017) "Using eDNA to Detect the Distribution and Density of Invasive Crayfish in the Honghe-Hani Rice Terrace World Heritage Site." PloS One 12 (5):e0177724.

Cannon, M. V., J. Hester, A. Shalkhauser, E. R. Chan, K. Logue, S. T. Small, and D. Serre. (2016) "In Silico Assessment of Primers for eDNA Studies Using PrimerTree and Application to

Characterize the Biodiversity Surrounding the Cuyahoga River." Scientific Reports 6 (March):22908.

Capella-Gutiérrez, S, Silla-Martínez, J.M. and Gabaldón.T. (2009) "trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." Bioinformatics 25 (15):1972–73.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K Noah Fierer. (2010) "QIIME Allows Analysis of High-Throughput Community Sequencing Data." Nature Methods 7 (5):335–36.

Carew, M.E.., Pettigrove, V.J., Metzeling L. and Hoffmann. A.A. (2013) "Environmental Monitoring Using next Generation Sequencing: Rapid Identification of Macroinvertebrate Bioindicator Species." Frontiers in Zoology 10 (1):45.

Catchen, J, Hohenlohe, P.A., Bassham, A., Amores, A., and Cresko. W.A., (2013) "Stacks: An Analysis Tool Set for Population Genomics." Molecular Ecology 22 (11):3124–40.

Chen, W, Bierbach, D., Plath, M., Streit, B. and Klaus. S. (2012) "Distribution of Amphipod Communities in the Middle to Upper Rhine and Five of Its Tributaries." BioInvasions Records 1 (4):263–71.

Chen, H. (2016). VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.6.17. https://CRAN.R-project.org/package=VennDiagram

Cock, P.J. Antao, A.T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., (2009) "Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics." Bioinformatics 25 (11):1422–23.

Coissac, E, Riaz, T. and Puillandre. N. (2012) “Bioinformatic Challenges for DNA Metabarcoding of Plants and Animals.” Molecular Ecology 21 (8):1834–47.

Connelly, N A., Charles R. O’Neill Jr, Barbara A. Knuth, and Tommy L. Brown. (2007) “Economic Impacts of Zebra Mussels on Drinking Water Treatment and Electric Power Generation Facilities.” Environmental Management 40 (1):105–12.

Constable, D, and Birkby. N.J. (2016) “The Impact of the Invasive Amphipod *Dikerogammarus Haemobaphes* on Leaf Litter Processing in UK Rivers.” Aquatic Ecology 50 (2). Springer Netherlands:273–81.

Copilaş-Ciocianu D, Grabowski M, Pârvulescu L, Petrusek A (2014) Zoogeography of epigean freshwater Amphipoda (Crustacea) in Romania: fragmented distributions and wide altitudinal variability. Zootaxa 3893: 243-260

Copilaş-Ciocianu, D, and Petrusek. A. (2015) “The Southwestern Carpathians as an Ancient Centre of Diversity of Freshwater Gammarid Amphipods: Insights from the *Gammarus Fossarum* Species Complex.” Molecular Ecology 24 (15):3980–92.

Creer, S, Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W.K., Potter, C. and Bik. H.M. (2016) “The Ecologist’s Field Guide to Sequence-Based Identification of Biodiversity.” Methods in Ecology and Evolution / British Ecological Society 7 (9):1008–18.

Darling, J A., and Mahon. A.R (2011) “From Molecules to Management: Adopting DNA-Based Methods for Monitoring Biological Invasions in Aquatic Environments.” Environmental Research 111 (7):978–88.

Davis, M A. (2009) Invasion Biology. Oxford University Press on Demand.

Davison, P. I., V. Créach, W-J Liang, D. Andreou, J. R. Britton, and G. H. Copp. 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 (3):1782–93.

Davy-Bowker, J., R. T. Clarke, T. A. Corbin, and H. M. Vincent. (2008) "River Invertebrate Classification Tool." eprints.bournemouth.ac.uk. http://eprints.bournemouth.ac.uk/16550/.

Davy, C.M., Kidd, A.G. and Wilson. C.C. (2015) "Development and Validation of Environmental DNA (eDNA) Markers for Detection of Freshwater Turtles." PloS One 10 (7):e0130965.

Deagle, B.E., Eveson, J.P. and Jarman. S.M. (2006) "Quantification of Damage in DNA Recovered from Highly Degraded Samples--a Case Study on DNA in Faeces." Frontiers in Zoology 3 (1). BioMed Central:11.

Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F. and Taberlet. P. (2014) "DNA Metabarcoding and the Cytochrome c Oxidase Subunit I Marker: Not a Perfect Match." Biology Letters 10 (9).

Deiner, K, and Altermatt. F. (2014) "Transport Distance of Invertebrate Environmental DNA in a Natural River." PloS One 9 (2):e88786.

Deiner, K, Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S. (2017) "Environmental DNA Metabarcoding: Transforming How We Survey Animal and Plant Communities." Molecular Ecology 26 (21):5872–95.

Deiner, K, Fronhofer, E., Mächler, E., Walser, J.C. and Altermatt. F. (2016) "Environmental DNA Reveals That Rivers Are Conveyer Belts of Biodiversity Information." Nature Communications 7 (August). Nature Publishing Group:12544.

Dejean, T, Valentini, A., Duparc, A., Pellier-Cuit, S., Pompanon, F., Taberlet, P. and Miaud. C. (2011) “Persistence of Environmental DNA in Freshwater Ecosystems.” PloS One 6 (8):e23398.

Dejean, T, Valentini, A, Miquel, C., Taberlet, P., Bellemain, E. and Miaud. C. (2012) “Improved Detection of an Alien Invasive Species through Environmental DNA Barcoding: The Example of the American Bullfrog Lithobates Catesbeianus.” The Journal of Applied Ecology 49 (4). Blackwell Publishing Ltd:953–59.

De Ventura, L, Kopp, K., Seppälä, K and 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.” mBio 8 (1):101–12.

Dick, J. T. A., D. Platvoet and Kelly. D.W. (2002) “Predatory Impact of the Freshwater Invader *Dikerogammarus Villosus* (Crustacea: Amphipoda).” Canadian Journal of Fisheries and Aquatic Sciences. Journal Canadien Des Sciences Halieutiques et Aquatiques 59 (6):1078–84.

Dick, J. T. A., and D. Platvoet. (2000) “Invading Predatory Crustacean *Dikerogammarus Villosus* Eliminates Both Native and Exotic Species.” Proceedings of the Royal Society of London B: Biological Sciences 267 (1447). The Royal Society:977–83.

Doi H, Uchii K, and Takahara T, (2015a) Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. PloS one 10(3): e0122763.

Doi, H., Takahara, T., Minamoto, T., Matsuhashi, S., Uchii, K. and Yamanaka, H. (2015b) “Droplet Digital Polymerase Chain Reaction (PCR) Outperforms Real-Time PCR in the Detection of Environmental DNA from an Invasive Fish Species.” Environmental Science & Technology 49 (9):5601–8.

Dougherty, M. M., Larson, E.R., Renshaw, M.A., Gantz, C.A., Egan, S.P., Erickson, D.E. and Lodge. D.M. (2016) "Environmental DNA (eDNA) Detects the Invasive Rusty Crayfish Orconectes Rusticus at Low Abundances." The Journal of Applied Ecology 53 (3):722–32.

Dunn, N, Priestley, V., Herraiz, A., Arnold, R. and Savolainen. V. (2017) "Behavior and Season Affect Crayfish Detection and Density Inference Using Environmental DNA." Ecology and Evolution 7 (19):7777–85.

Eddy, S.R. (1998) Profile Hidden Markov Models. Bioinformatics, 14:755-763.

Egan, S P., Barnes, M.A., Hwang, C.T., Mahon, A.R., Feder, J.L., Ruggiero, S.T., Tanner, C.E. and Lodge. M. (2013) "Rapid Invasive Species Detection by Combining Environmental DNA with Light Transmission Spectroscopy." Conservation Letters 6 (6):402–9.

Eggers TO, Martens A (2001) A key to the freshwater Amphipoda (Crustacea) of Germany. Lauterbornia 55: 1-68

Eisenring, M, F. Altermatt, A.M. Westram, and J. Jokela. (2016) "Habitat Requirements and Ecological Niche of Two Cryptic Amphipod Species at Landscape and Local Scales." Ecosphere 7 (5).

Elbrecht, V, and Leese. F. (2015) "Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass--Sequence Relationships with an Innovative Metabarcoding Protocol." PloS One 10 (7): e0130324.

Elbrecht, V, and Leese. F. (2016) "Validation and Development of Freshwater Invertebrate Metabarcoding COI Primers for Environmental Impact Assessment." e2044v3. PeerJ Preprints.

Elbrecht, V, B. Peinert, and F. Leese. (2017) “Sorting Things out - Assessing Effects of Unequal Specimen Biomass on DNA Metabarcoding.” e2561v3. PeerJ Preprints.

Environment Agency (2017) Operational instruction 018_08 Freshwater macroinvertebrate sampling in rivers. Environment Agency, Bristol, UK

EU-STAR (2004) UK Invertebrate sampling and analysis procedure for STAR project, RIVPACS Macro-invertebrate sampling protocol

Evans, N. T., Olds, B.P., Renshaw, M.A., Turner, C.R., Li, Y., Jerde, C.L. Mahon, A.R., Pfrender, M.E.., Lamberti, G.A. and Lodge. D.M. (2016) “Quantification of Mesocosm Fish and Amphibian Species Diversity via Environmental DNA Metabarcoding.” Molecular Ecology Resources 16 (1):29–41.

Feckler, A., A. Thielsch, K. Schwenk, R. Schulz, and M. Bundschuh. (2012) “Differences in the Sensitivity among Cryptic Lineages of the *Gammarus Fossarum* Complex.” The Science of the Total Environment 439 (November):158–64.

Ficetola, G.F., C. Miaud, F. Pompanon, and P. Taberlet. (2008) “Species Detection Using Environmental DNA from Water Samples.” Biology Letters 4 (4):423–25.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S., Heger, A. et al. (2014) “Pfam: The Protein Families Database.” Nucleic Acids Research 42 (Database issue): D222–30.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. (1994) “DNA Primers for Amplification of Mitochondrial Cytochrome c Oxidase Subunit I from Diverse Metazoan Invertebrates.” Molecular Marine Biology and Biotechnology 3 (5):294–99.

Foote, A. D., Thomsen, P., Sveegaard, S., Wahlberg, M., Kielgast, J., Kyhn, L.A., Salling, A.B., Galatius, A., Orlando, L. and Gilbert. (2012) M "Investigating the Potential Use of Environmental DNA (eDNA) for Genetic Monitoring of Marine Mammals." PloS One 7 (8):e41781.

Furse, M. T., J. F. Wright, P. D. Armitage, and D. Moss. (1981) "An Appraisal of Pond-Net Samples for Biological Monitoring of Lotic Macro-Invertebrates." Water Research 15 (6):679–89.

Gallardo, B, and D.C. Aldridge. (2013a) "Priority Setting for Invasive Species Management: Risk Assessment of Ponto-Caspian Invasive Species into Great Britain." Ecological Applications: A

Gallardo, B, and D.C. Aldridge. (2013b) "The 'dirty Dozen': Socio-Economic Factors Amplify the Invasion Potential of 12 High-Risk Aquatic Invasive Species in Great Britain and Ireland." The Journal of Applied Ecology 50 (3):757–66.

Gallardo, B, and D.C. Aldridge. (2014) "Is Great Britain Heading for a Ponto–Caspian Invasional Meltdown?" The Journal of Applied Ecology. https://doi.org/10.1111/1365-2664.12348.

Goedmakers A (1972) *Gammarus fossarum* Koch, 1835: Redescription based on neotype material and notes on its local variation (Crustacea, Amphipoda). Bijdragen Tot de Dierkunde 42: 124-138

Goldberg, C S., Sepulveda, A., Ray, A., Baumgardt, J. and Waits. L.P. (2013) "Environmental DNA as a New Method for Early Detection of New Zealand Mudsnails (Potamopyrgus Antipodarum)." Freshwater Science 32 (3). The University of Chicago Press on behalf of Society for Freshwater Science:792–800.

Goldberg, C S., Strickler, K.M. and Pilliod. D.S. (2015) “Moving Environmental DNA Methods from Concept to Practice for Monitoring Aquatic Macroorganisms.” Biological Conservation 183 (March):1–3.

Goldberg, C S., Turner, C.R., Deiner, K., Klymus, K.E., Thomsen, P.F., Murphy, M.A., Spear, S.E. (2016) “Critical Considerations for the Application of Environmental DNA Methods to Detect Aquatic Species.” Methods in Ecology and Evolution / British Ecological Society, July.

Grabowski, M. (2007) “Alien Crustacea in Polish Waters – Amphipoda.” Aquatic Invasions 2 (1):25–38.

Greninger, A., L., Samia N. Naccache, Scot Federman, Guixia Yu, Placide Mbala, Vanessa Bres, Doug Stryke, et al. (2015) “Rapid Metagenomic Identification of Viral Pathogens in Clinical Samples by Real-Time Nanopore Sequencing Analysis.” Genome Medicine 7 (1):99.

Gweon, H.S. (2017) PIPCOI pipeline. Centre of Ecology and Hydrology, United Kingdom. unpublished data

Haase, P, Murray-Bligh, J., Lohse, S., Pauls, S., Sundermann, A., Gunn, R. and Clarke. R. (2006) “Assessing the Impact of Errors in Sorting and Identifying Macroinvertebrate Samples.” Hydrobiologia 566 (1):505–21.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.C. and Baird. D.J. (2011) “Environmental Barcoding: A next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos.” PloS One 6 (4): e17497.

Hänfling, B, Edwards, F. and Gherardi. F. (2011) “Invasive Alien Crustacea: Dispersal, Establishment, Impact and Control.” Biocontrol 56 (4). Springer Netherlands:573–95.

Hänfling, B., Lawson Handley, L., Read, D.S., Hahn, C., Li, J., Nichols, P., Blackman, R.C., Oliver, A., and Winfield. I. (2016) "Environmental DNA Metabarcoding of Lake Fish Communities Reflects Long-Term Data from Established Survey Methods." Molecular Ecology 25 (13):3101–19.

Hänfling, B., Lawson Handley, L., Lunt, D., Shum, P., Winfield, I and Read, D. (2017) A Review of Recent Advances in Genetic Methods to Identify Improvements in CAMERAS Partners Monitoring Activities. Report to the Scottish Government, 2017. REFERENCE – CR/2015/09

Harper, LR., Lawson Handley, L., Hahn, C., Boonham. N., Rees, HC., Gough, KC., Ian, EL., Adams, P., Brotherton, P., Philips, S and Hänfling, B. (2017) Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of great crested newt (Triturus cristatus) BioRxiv 215897

Hou, Z., Sket, B., Fiser, C. and Li, S. (2011) "Eocene Habitat Shift from Saline to Freshwater Promoted Tethyan Amphipod Diversification." Proceedings of the National Academy of Sciences of the United States of America 108 (35):14533–38.

Howald, G.D., Galván, C.J., Russell, J.P., Parkes, J.C., Samaniego, J., Wang, Y. (2007) "Invasive Rodent Eradication on Islands." Conservation Biology: The Journal of the Society for Conservation Biology 21 (5):1258–68.

Hoy, M.S., Kelly, K. and Rodriguez, R.J. (2010) "Development of a Molecular Diagnostic System to Discriminate *Dreissena Polymorpha* (zebra Mussel) and Dreissena Bugensis (quagga Mussel)." Molecular Ecology Resources 10 (1):190–92.

Hulme, P.E. (2006) "Beyond Control: Wider Implications for the Management of Biological Invasions." The Journal of Applied Ecology 43 (5). Blackwell Publishing Ltd:835–47.

Hulme, P.E. (2009) “Trade, Transport and Trouble: Managing Invasive Species Pathways in an Era of Globalization.” The Journal of Applied Ecology 46 (1). Blackwell Publishing Ltd:10–18.

Huson, D.H., Auch, A.F., Qi, J. and Schuster. S.C. (2007) “MEGAN Analysis of Metagenomic Data.” Genome Research 17 (3):377–86.

Jackson, J.K., Battle, J.M., White, B.P., Pilgrim, E.M., Stein, E.D., Miller, P.E. and Sweeney. B.W., (2014) “Cryptic Biodiversity in Streams: A Comparison of Macroinvertebrate Communities Based on Morphological and DNA Barcode Identifications.” Freshwater Science 33 (1). [The University of Chicago Press, Society for Freshwater Science]:312–24.

Jane, S. F., Wilcox, T.M., McKelvey, K.S., Young, M.K., Schwartz, M.K., Lowe, W.H., Letcher, B.H. and Whiteley. A.R. (2015) “Distance, Flow and PCR Inhibition: eDNA Dynamics in Two Headwater Streams.” Molecular Ecology Resources 15 (1):216–27.

Janetzky, W. (1994) “Distribution of the Genus *Gammarus* (Amphipoda: Gammaridae) in the River Hunte and Its Tributaries (Lower Saxony, Northern Germany).” Hydrobiologia. Springer.

Jerde, C. L., Mahon, A.M., Chadderton, W.L. and Lodge. D.M. (2011) “‘Sight-Unseen’ Detection of Rare Aquatic Species Using Environmental DNA.” Conservation Letters 4 (2). Blackwell Publishing Inc:150–57.

Jerde, C.L., Olds, B.P., Shogren, A.J., Andruszkiewicz, E.A., Mahon, A.R., Bolster, D. and Tank. J.L. (2016) “Influence of Stream Bottom Substrate on Retention and Transport of Vertebrate Environmental DNA.” Environmental Science & Technology 50 (16):8770–79.

Karaman GS, Pinkster S (1977) Freshwater *Gammarus* species from Europe, North Africa and adjacent regions of Asia (Crustacea, Amphipoda). Part I. *Gammarus pulex*-group and related species. Bijdragen Tot de Dierkunde 47: 1-97

Karatayev, A.Y., Burlakova, L.E. and Padilla. D.K (2002) “Impacts of Zebra Mussels on Aquatic Communities and Their Role as Ecosystem Engineers.” In Invasive Aquatic Species of Europe. Distribution, Impacts and Management, 433–46. Springer Netherlands.

Karatayev, A.Y., Padilla, D.K., Minchin, D., Boltovskoy, D. and Burlakova. L.E. (2007) “Changes in Global Economies and Trade: The Potential Spread of Exotic Freshwater Bivalves.” Biological Invasions 9 (2). Kluwer Academic Publishers:161–80.

Katoh, K. and Daron M. Standley. (2013) “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.” Molecular Biology and Evolution 30 (4):772–80.

Kelly, R P., Port, J.A., Yamahara, K.M and Crowder. L.B. (2014) “Using Environmental DNA to Census Marine Fishes in a Large Mesocosm.” PloS One 9 (1):e86175.

Kinzler, W., Gerd Mayer, A.K., Waloszek, D. and Maier. G. (2009) “Mutual Predation between and Cannibalism within Several Freshwater Gammarids: *Dikerogammarus Villosus* versus One Native and Three Invasives.” Aquatic Ecology 43 (2). Springer Netherlands:457.

Kitson JJN, Hahn C, Sands RJ, Straw NA, Evans DM, Lunt DH (2015) Nested metabarcode tagging: a robust tool for studying species interactions in ecology and evolution. bioRxiv 035071

Klymus, K., Richter, C., Chapman, D.C., Paukert. C. (2014) “Quantification of eDNA Shedding Rates from Invasive Bighead Carp Hypophthalmichthys Nobilis and Silver Carp Hypophthalmichthys Molitrix.” Biological Conservation

Klymus, K.E., Marshall, N.T. and Stepien. C.A. (2017) “Environmental DNA (eDNA) Metabarcoding Assays to Detect Invasive Invertebrate Species in the Great Lakes.” PloS One 12 (5):e0177643.

Klymus, K.E., Richter, C.A., Chapman, D.C. and Paukert. C. (2015) “Quantification of eDNA Shedding Rates from Invasive Bighead Carp Hypophthalmichthys Nobilis and Silver Carp Hypophthalmichthys Molitrix.” Biological Conservation 183 (March):77–84.

Lawson Handley, L. (2015) “How Will the ‘molecular Revolution’ Contribute to Biological Recording?” Biological Journal of the Linnean Society. Linnean Society of London 115 (3):750–66.

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, 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.

Leray, M, Joy Y. Yang, Christopher P. Meyer, Suzanne C. Mills, Natalia Agudelo, Vincent Ranwez, Joel T. Boehm, and Ryuji J. Machida. (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 (June):34.

Lodge, D.M., Williams, S., MacIsaac, H.J., Hayes, K.R., Leung, B., Reichard, S., Mack, R.N. (2006) "Biological Invasions: Recommendations for U.S. Policy and Management." Ecological Applications: A Publication of the Ecological Society of America 16 (6):2035–54.

Macher, J, and Leese. F. (2017) "Environmental DNA Metabarcoding of Rivers: Not All eDNA Is Everywhere, and Not All the Time." bioRxiv. https://doi.org/10.1101/164046.

Mächler, E., Deiner, K., Steinmann, P. and Altermatt. F. (2014) "Utility of Environmental DNA for Monitoring Rare and Indicator Macroinvertebrate Species." Freshwater Science 33 (4):1174–83.

MacKenzie, D. I., Nichols, J.D., Hines, J.E., Knutson, M.G. and Franklin. A.B. (2003) "Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly." Ecology 84 (8). Ecological Society of America:2200–2207.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, A. and Langtimm. C.A. (2002) "Estimating site occupancy rates when detection probabilities are less than one." Ecology 83 (8). Ecological Society of America:2248–55.

MacKenzie DI and Nichols JD (2004) Occupancy as a surrogate for abundance estimation. Animal: an international journal of animal bioscience, abc.museucienciesjournals.cat.

MacNeil C, Platvoet D (2005) The predatory impact of the freshwater invader *Dikerogammarus villosus* on native *Gammarus pulex* (Crustacea: Amphipoda); influences of differential microdistribution and food resources. Journal of Zoology 267: 31-38

MacNeil, C, P. Boets, and D. Platvoet. (2012) "'Killer Shrimps', Dangerous Experiments and Misguided Introductions; How Freshwater Shrimp (Crustacea: Amphipoda) Invasions Threaten Biological Water Quality Monitoring in the British Isles." Freshwater Reviews: A Journal of the Freshwater Biological Association 5 (1):21–35.

MacNeil, C, D. Platvoet, J. Dick, N. Fielding, A. Constable, N. Hall, D. Aldridge, T. Renals, and M. Diamond. (2010a) "The Ponto-Caspian 'killer Shrimp', *Dikerogammarus Villosus* (Sowinsky, 1894), Invades the British Isles." Aquatic Invasions 5 (4):441–45.

MacNeil, C, J T. A. Dick, D Platvoet, and M Briffa. (2010b) "Direct and Indirect Effects of Species Displacements: An Invading Freshwater Amphipod Can Disrupt Leaf-Litter Processing and Shredder Efficiency." Journal of the North American Benthological Society 30 (1). The Society for Freshwater Science:38–48.

Magoč, T., and Salzberg, S.L. (2011) "FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies." Bioinformatics 27 (21):2957–63.

Mahon, A.R., M.A. Barnes, S. Senapati, J.L Feder, J.A. Darling, H. Chang, and D.M. Lodge. (2011) "Molecular Detection of Invasive Species in Heterogeneous Mixtures Using a Microfluidic Carbon Nanotube Platform." PloS One 6 (2).

Mahon, A.R., L.R. Nathan, and C.L Jerde. (2014) "Meta-Genomic Surveillance of Invasive Species in the Bait Trade." Conservation Genetics Resources 6.

Manchester, S. J., and J.M. Bullock. (2000) "The Impacts of Non-Native Species on UK Biodiversity and the Effectiveness of Control." The Journal of Applied Ecology 37 (5). Blackwell Science Ltd:845–64.

Martens A, Grabow K (2008) Das risiko der verschleppung neozoischer Amphipoda beim überlandtransport von yachten. Lauterbornia 62: 41–44

Meijering, M.P.D. (1971) Die Gammarus-Fauna der Schlitzerländer Fließgewässer. Archiv für Hydrobiologie 68: 575-608

Meijering, M.P.D. (1972) Physiologische Beiträge zur Frage der systematischen Stellung von *Gammarus pulex* (L.) und *Gammarus fossarum* Koch (Amphipoda). Crustaceana Suppl., 3: 313-325

Metcalfe, J. L. (1989) "Biological Water Quality Assessment of Running Waters Based on Macroinvertebrate Communities: History and Present Status in Europe." Environmental Pollution 60 (1-2):101–39.

Miya, M., Y. Sato, T. Fukunaga, T. Sado, J. Y. Poulsen, K. Sato, T. Minamoto, et al. (2015) "MiFish, a Set of Universal PCR Primers for Metabarcoding Environmental DNA from Fishes: Detection of More than 230 Subtropical Marine Species." Royal Society Open Science 2 (7). The Royal Society:150088.

Murray-Bligh J.A.D. (1999) Procedure for Collecting and Analysing Macroinvertebrate Samples, Environment Agency, Bristol, UK

Nalepa, T. F., D.L. Fanslow, and S. A. Pothoven. (2010) "Recent Changes in Density, Biomass, Recruitment, Size Structure, and Nutritional State of Dreissena Populations in Southern Lake Michigan." Journal of Great Lakes Research 36 (January):5–19.

Nathan, L.M., M. Simmons, B.J. Wegleitner, C.L. Jerde, and A.R. Mahon. (2014) "Quantifying Environmental DNA Signals for Aquatic Invasive Species across Multiple Detection Platforms." Environmental Science & Technology 48 (21):12800–806.

Nathan, L.R., C.L. Jerde, M.L. Budny, and A.R. Mahon. (2015) "The Use of Environmental DNA in Invasive Species Surveillance of the Great Lakes Commercial Bait Trade." Conservation Biology: The Journal of the Society for Conservation Biology 29 (2):430–39.

Nijssen, H., (1963) Some notes on the distribution and the ecology of the amphipod *Gammarus fossarum* Koch, 1835, in the Netherlands (Crustacea, Malacostraca). Beaufortia, 10 (116): 40—43.

Oksanen. J, Guillaume Blanchet. F, Friendly. M, Kindt, R., Legendre, P, McGlinn, D, Minchin, P.R, O'Hara, R.B, Simpson, G.L, Solymos, P, Henry M, Stevens, H, Szoecs, E and Wagner H (2017). vegan: Community Ecology Package. R package version 2.4-3.https://CRAN.R-project.org/package=vegan

Olsen, J. B., C.J. Lewis, R.L. Massengill, K.J. Dunker, and J.K Wenburg. (2015) "An Evaluation of Target Specificity and Sensitivity of Three qPCR Assays for Detecting Environmental DNA from Northern Pike (Esox Lucius)." Conservation Genetics Resources 7 (3).

Oreska, M.P.J., and Aldridge, D.C. (2011) "Estimating the Financial Costs of Freshwater Invasive Species in Great Britain: A Standardized Approach to Invasive Species Costing." Biological Invasions 13 (2). Springer Netherlands:305–19.

Paula, D. P., B. Linard, D.A. Andow, E.R. Sujii, C.S.S. Pires, and A.P. Vogler. (2015) "Detection and Decay Rates of Prey and Prey Symbionts in the Gut of a Predator through Metagenomics." Molecular Ecology Resources 15 (4):880–92.

Paula, D P., Linard, B., Crampton-Platt, A., Srivathsan, A., Timmermans, M.J.T.N., Sujii, E.R., Pires, C.S.S., Souza, L.M., Andow, D.A. and Vogler. A.P. 2016. "Uncovering Trophic Interactions in Arthropod Predators through DNA Shotgun-Sequencing of Gut Contents." PloS One 11 (9):e0161841.

Peeters ETHM, Gardeniers JJP (1998) Logistic regression as a tool for defining habitat requirements of two common gammarids. Freshwater Biology 39: 605–615.

Peñarrubia, L, C. Alcaraz, A. Bij de Vaate, N. Sanz, C. Pla, O. Vidal, and J. Viñas. (2016) "Validated Methodology for Quantifying Infestation Levels of Dreissenid Mussels in Environmental DNA (eDNA) Samples." Scientific Reports 6 (December):39067.

Peyer, S. M., A.J. McCarthy, and C.E. Lee. (2009) "Zebra Mussels Anchor Byssal Threads Faster and Tighter than Quagga Mussels in Flow." The Journal of Experimental Biology 212 (Pt 13):2027–36.

Pilliod, D. S., C.S. Goldberg, R.S. Arkle, and L.P. Waits. (2013) "Estimating Occupancy and Abundance of Stream Amphibians Using Environmental DNA from Filtered Water Samples." Canadian Journal of Fisheries and Aquatic Sciences. Journal Canadien Des Sciences Halieutiques et Aquatiques 70 (8):1123–30.

Pinkster S (1972) On members of the *Gammarus pulex*-group (Crustacea-Amphipoda) from western Europe. Bijdragen Tot de Dierkunde 42: 164-191

Piscart C, Mermillod-Blondin F, Maazouzi C, Merigoux S, Marmonier P (2011) Potential impact of invasive amphipods on leaf litter recycling in aquatic ecosystems. Biological Invasions 13: 2861-2868

Piscart C, Bollache L (2012) Crustacés amphipodes de surface (Gammare d'eau douce). Association Française de Limnologie, Thonon les Bains, 113pp

Port, J.A., J.L. O'Donnell, O.C. Romero-Maraccini, P.R. Leary, S.Y. Litvin, K.J. Nickols, K.M. Yamahara, and R.P. Kelly. (2016) "Assessing Vertebrate Biodiversity in a Kelp Forest Ecosystem Using Environmental DNA." Molecular Ecology 25 (2):527–41.

Quick, J, N. J. Loman, S. Duraffour, J.T. Simpson, E. Severi, L. Cowley, J.A. Bore, et al. (2016) "Real-Time, Portable Genome Sequencing for Ebola Surveillance." Nature 530 (7589):228–32.

Quinn, A, B. Gallardo, and D.C. Aldridge. (2014) "Quantifying the Ecological Niche Overlap between Two Interacting Invasive Species: The Zebra Mussel (*Dreissena Polymorpha*) and the Quagga Mussel (*Dreissena Rostriformis Bugensis*)." Aquatic Conservation: Marine and Freshwater Ecosystems 24 (3):324–37.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Radulovici, A. E., B. Sainte-Marie, and F. Dufresne. (2009) "DNA Barcoding of Marine Crustaceans from the Estuary and Gulf of St Lawrence: A Regional-Scale Approach." Molecular Ecology Resources 9 Suppl s1 (May):181–87.

Rees, H.C., B. C. Maddison, D.J., Middleditch, J.R.M. Patmore, and K.C. Gough. (2014) "REVIEW: The Detection of Aquatic Animal Species Using Environmental DNA – a Review of eDNA as a Survey Tool in Ecology." The Journal of Applied Ecology 51 (5):1450–59.

Rho, M., Tang, H. and Ye, Y. (2010) "FragGeneScan: Predicting Genes in Short and Error-Prone Reads." Nucleic Acids Research 38 (20):e191.

Ricciardi, A, R.J. Neves, and J.B. Rasmussen. (1998) "Impending Extinctions of North American Freshwater Mussels (Unionoida) Following the Zebra Mussel (*Dreissena Polymorpha*) Invasion." The Journal of Animal Ecology 67 (4). Blackwell Science Ltd:613–19.

Ricciardi, A., R. Serrouya, and F. G. Whoriskey. (1995) “Aerial Exposure Tolerance off Zebra and Quagga Mussels (Bivalvia: Dreissenidae): Implications for Overland Dispersal.” Canadian Journal of Fisheries and Aquatic Sciences. Journal Canadien Des Sciences Halieutiques et Aquatiques 52 (3). NRC Research Press:470–77.

Ricciardi, A., F. G. Whoriskey, and J. B. Rasmussen. (1996) “Impact of the(Dreissena) Invasion on Native Unionid Bivalves in the Upper St. Lawrence River.” Canadian Journal of Fisheries and Aquatic Sciences. Journal Canadien Des Sciences Halieutiques et Aquatiques 53 (6). NRC Research Press:1434–44.

Roe, S. L., and H. J. MacIsaac. (1997) “Deepwater Population Structure and Reproductive State of Quagga Mussels (Dreissena Bugensis) in Lake Erie.” Canadian Journal of Fisheries and Aquatic Sciences. Journal Canadien Des Sciences Halieutiques et Aquatiques 54 (10):2428–33.

Rognes, T, T. Flouri, B. Nichols, C. Quince, and F. Mahé. (2016) “VSEARCH: A Versatile Open Source Tool for Metagenomics.” e2409v1. PeerJ Preprints.

Roussel, J., J. Paillisson, A. Tréguier, and E. Petit. (2015) “The Downside of eDNA as a Survey Tool in Water Bodies.” The Journal of Applied Ecology 52 (4):823–26.

Roy, H.E., J. Peyton, D.C. Aldridge, T. Bantock, T.M. Blackburn, R. Britton, P. Clark (2014) “Horizon Scanning for Invasive Alien Species with the Potential to Threaten Biodiversity in Great Britain.” Global Change Biology 20 (12):3859–71.

Sansom, B.J., and L.M. Sassoubre. (2017) “Environmental DNA (eDNA) Shedding and Decay Rates to Model Freshwater Mussel eDNA Transport in a River.” Environmental Science & Technology,

Sassoubre, Lauren M., Kevan M. Yamahara, Luke D. Gardner, Barbara A. Block, and Alexandria B. Boehm. (2016) "Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine Fish." Environmental Science & Technology, September.

Schmelzle, M.C., and Kinziger. A.P. (2016) "Using Occupancy Modelling to Compare Environmental DNA to Traditional Field Methods for Regional-Scale Monitoring of an Endangered Aquatic Species." Molecular Ecology Resources 16 (4):895–908.

Schneider, J., Valentini, A., Dejean, T., Montarsi, F., Taberlet, P., Glaizot, O. and Fumagalli. L. (2016) "Detection of Invasive Mosquito Vectors Using Environmental DNA (eDNA) from Water Samples." PloS One 11 (9):e0162493.

Scriver, M, Marinich, A., Wilson, C. and Freeland. J. (2015) "Development of Species-Specific Environmental DNA (eDNA) Markers for Invasive Aquatic Plants." Aquatic Botany 122:27–31.

Shogren, A. J., Tank, J.L., Andruszkiewicz, E., Olds, B., Mahon, A.R., Jerde, C.L. and Bolster. D. (2017) "Controls on eDNA Movement in Streams: Transport, Retention, and Resuspension." Scientific Reports 7 (1):5065.

Shokralla, S., Gibson, J., King, I., Baird, D., Janzen, D., Hallwachs, W. and Hajibabaei. M. (2016) "Environmental DNA Barcode Sequence Capture: Targeted, PCR-Free Sequence Capture for Biodiversity Analysis from Bulk Environmental Samples." bioRxiv.

Shokralla, S, Spall, J.L., Gibson, J.F. and Hajibabaei. M. (2012) "Next-Generation Sequencing Technologies for Environmental DNA Research." Molecular Ecology 21 (8):1794–1805.

Simmons, M., Tucker, A., Chadderton, L.W., Jerde, C.L. and Mahon. A.R. (2015) "Active and Passive Environmental DNA Surveillance of Aquatic Invasive Species." Canadian Journal of

Fisheries and Aquatic Sciences. Journal Canadien Des Sciences Halieutiques et Aquatiques 73 (1). NRC Research Press:76–83.

Smart, A.S., R. Tingley, A.R. Weeks, A. R. van Rooyen, and M.A. McCarthy. (2015) "Environmental DNA Sampling Is More Sensitive than a Traditional Survey Technique for Detecting an Aquatic Invader." Ecological Applications: A Publication of the Ecological Society of America 25 (7). Ecological Society of America:1944–52.

Stamatakis, A. (2014) "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." Bioinformatics 30 (9):1312–13.

Stoeckle, B. C., R. Kuehn, and J.Geist. (2016) "Environmental DNA as a Monitoring Tool for the Endangered Freshwater Pearl Mussel (Margaritifera Margaritifera L.): A Substitute for Classical Monitoring Approaches?" Aquatic Conservation: Marine and Freshwater Ecosystems 26 (6):1120–29.

Stoeckle, M. Y., L. Soboleva, and Z. Charlop-Powers. (2017) "Aquatic Environmental DNA Detects Seasonal Fish Abundance and Habitat Preference in an Urban Estuary." PloS One 12 (4):e0175186.

Strayer, D. L., and D. Dudgeon. (2010) "Freshwater Biodiversity Conservation: Recent Progress and Future Challenges." Journal of the North American Benthological Society 29 (1). The Society for Freshwater Science:344–58.

Strickler, K.M., A.K. Fremier, and C.S. Goldberg. (2015) "Quantifying Effects of UV-B, Temperature, and pH on eDNA Degradation in Aquatic Microcosms." Biological Conservation 183 (March):85–92.

Sutherland, W.J., S.Bardsley, M.Clout, M.H. Depledge, L.V. Dicks, L. Fellman, E.Fleishman. (2013) “A Horizon Scan of Global Conservation Issues for 2013.” Trends in Ecology & Evolution 28 (1). Elsevier:16–22.

Sweeney, B.W., Battle, J.M., Jackson, J.K. and Dapkey. T. (2011) “Can DNA Barcodes of Stream Macroinvertebrates Improve Descriptions of Community Structure and Water Quality?” Journal of the North American Benthological Society 30 (1):195–216.

Szitenberg, A, John, M., Blaxter, M.L. and Lunt. D.H. (2015) “ReproPhylo: An Environment for Reproducible Phylogenomics.” PLoS Computational Biology 11 (9). Public Library of Science:e1004447.

Taberlet, P., Coissac, E., Hajibabaei, M. and Rieseberg. L.H. (2012a) “Environmental DNA.” Molecular Ecology 21 (8):1789–93.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. and Willerslev. E. (2012b) “Towards next-Generation Biodiversity Assessment Using DNA Metabarcoding.” Molecular Ecology 21 (8):2045–50.

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H. and Zen ’ichiro Kawabata. (2012) “Estimation of Fish Biomass Using Environmental DNA.” PloS One 7 (4):e35868.

Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G. and Fontaneto. D. (2012) “The Widely Used Small Subunit 18S rDNA Molecule Greatly Underestimates True Diversity in Biodiversity Surveys of the Meiofauna.” Proceedings of the National Academy of Sciences of the United States of America 109 (40):16208–12.

Thomsen, P.F., Kielgast, J., Lønsmann Iversen, L., Rask Møller, P., Rasmussen, M. and Willerslev. E. (2012a) "Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples." PloS One 7 (8). Public Library of Science: e41732.

Thomsen, P.F., Kielgast, J., Lønsmann Iversen, L., Wiuf, C., Rasmussen, M., Gilbert, TP., Orlando, L. and Willerslev. E. (2012b) "Monitoring Endangered Freshwater Biodiversity Using Environmental DNA." Molecular Ecology 21 (11):2565–73.

Thomsen, P.E., and Willerslev. E. (2015) "Environmental DNA – An Emerging Tool in Conservation for Monitoring Past and Present Biodiversity." Biological Conservation 183 (0):4–18.

Timar, L., and Phaneuf, D.J. (2009) "Modeling the Human-Induced Spread of an Aquatic Invasive: The Case of the Zebra Mussel." Ecological Economics: The Journal of the International Society for Ecological Economics 68 (12). Elsevier:3060–71.

Trebitz, A.S., Hoffman, J.C., Darling, J.A., Pilgrim, E.M., Kelly, J.R., Brown, E.A., Chadderton, W.L. (2017) "Early Detection Monitoring for Aquatic Non-Indigenous Species: Optimizing Surveillance, Incorporating Advanced Technologies, and Identifying Research Needs." Journal of Environmental Management 202 (Pt 1):299–310.

Tréguier, A., Paillisson, J., Dejean, T., Valentini, A., Schlaepfer, M.A. and Roussel. J. (2014) "Environmental DNA Surveillance for Invertebrate Species: Advantages and Technical Limitations to Detect Invasive Crayfish Procambarus Clarkii in Freshwater Ponds." The Journal of Applied Ecology 51 (4):871–79.

Turner, C.R., M.A. Barnes, C. C. Y. Xu, S.E. Jones, C.L. Jerde, and D.M. Lodge. (2014) "Particle Size Distribution and Optimal Capture of Aqueous Macrobial eDNA." bioRxiv.

Turner, C.R., D.J. Miller, K.J. Coyne, and J. Corush. (2014) “Improved Methods for Capture, Extraction, and Quantitative Assay of Environmental DNA from Asian Bigheaded Carp (Hypophthalmichthys Spp.).” PloS One 9 (12): e114329.

Turner, C.R., K.L. Uy, and R.C. Everhart. (2015) “Fish Environmental DNA Is More Concentrated in Aquatic Sediments than Surface Water.” Biological Conservation 183 (Supplement C):93–102.

Ushio, M., H. Fukuda, T. Inoue, K. Makoto, O.Kishida, K.Sato, K.Murata. (2017) “Environmental DNA Enables Detection of Terrestrial Mammals from Forest Pond Water.” Molecular Ecology Resources,

Valentini, A., Pompanon, F. and Taberlet. P. (2009) “DNA Barcoding for Ecologists.” Trends in Ecology & Evolution 24 (2):110–17.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E. (2016) “Next-Generation Monitoring of Aquatic Biodiversity Using Environmental DNA Metabarcoding.” Molecular Ecology, October. https://doi.org/10.1111/mec.13428.

Vamos, E.E., Elbrecht, V. and Leese. F. (2017) “Short COI Markers for Freshwater Macroinvertebrate Metabarcoding.” e3037v2. PeerJ Preprints.

Vander Zanden, M.J., Hansen, G.J.A., Higgins, S.N. and Kornis. M. (2010) “A Pound of Prevention, plus a Pound of Cure: Early Detection and Eradication of Invasive Species in the Laurentian Great Lakes.” Journal of Great Lakes Research 36 (1):199–205.

Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth edition. Springer.

Vivien, R., Wyler. S., Lafont, M. and Pawlowski. J. (2015) “Molecular Barcoding of Aquatic Oligochaetes: Implications for Biomonitoring.” PloS One 10 (4):e0125485.

Votintseva, A. A., Bradley, P., Pankhurst, L., Del Ojo Elias, C., Loose, M., Nilgiriwala, K., Chatterjee, A. (2017) “Same-Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct Respiratory Samples.” Journal of Clinical Microbiology 55 (5):1285–98.

Weiss, M. and Leese, F. (2016) “Widely Distributed and Regionally Isolated! Drivers of Genetic Structure in *Gammarus Fossarum* in a Human-Impacted Landscape.” BMC Evolutionary Biology 16 (July):153.

Weiss, M., Macher, J., Seefeldt, M.A. and Leese. F. (2014) “Molecular Evidence for Further Overlooked Species within the *Gammarus Fossarum*.” Hydrobiologia 721 (1). Springer Netherlands:165–84.

Wickham. H. (2009) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York

Wilcox, T.M., McKelvey, K.S., Young, M.K., Jane, S., Lowe, W.H., Whiteley, A.R. and Schwartz. M.K. (2013) “Robust Detection of Rare Species Using Environmental DNA: The Importance of Primer Specificity.” PloS One 8 (3):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. and Schwartz. M.K. (2016) “Understanding Environmental DNA Detection Probabilities: A Case Study Using a Stream-Dwelling Char Salvelinus Fontinalis.” Biological Conservation 194 (Supplement C):209–16.

Willerslev, E, Hansen, A.J., Gilbert, M.T.P. and Shapiro. B. (2003) “Diverse Plantand Animal Genetic Recordsfrom Holoceneand Pleistocene Sediments.” Science 300. (5620):791-5

Wurzbacher, C., Attermeyer, K., Kettner, M.T., Flintrop, C., Warthmann, N., Hilt, S., Grossart, H. and Monaghan. M.T. (2017) "DNA Metabarcoding of Unfractionated Water Samples Relates Phyto-, Zoo- and Bacterioplankton Dynamics and Reveals a Single-Taxon Bacterial Bloom." Environmental Microbiology Reports 9 (4):383–88.

WWF. 2016. Living Planet Report (2016) Risk and Resilience in a new era. WWF International, Gland, Switzerland

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden. T.L. (2012) "Primer-BLAST: A Tool to Design Target-Specific Primers for Polymerase Chain Reaction." BMC Bioinformatics 13 (June):134.

Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C. and Ding, Z. (2012) "Biodiversity Soup: Metabarcoding of Arthropods for Rapid Biodiversity Assessment and Biomonitoring." Methods in Ecology and Evolution / British Ecological Society 3 (4). Blackwell Publishing Ltd:613–23.

Zaiko, A., Martinez, J.L., Schmidt-Petersen, J., Ribicic, D., Samuiloviene, A. and Garcia-Vazquez. E., (2015) "Metabarcoding Approach for the Ballast Water Surveillance--an Advantageous Solution or an Awkward Challenge?" Marine Pollution Bulletin 92 (1-2):25–34.

Zhang, J., Kobert, K., Flouri, T. and Stamatakis. A. (2014) "PEAR: A Fast and Accurate Illumina Paired-End reAd mergeR." Bioinformatics 30 (5):614–20.

Zhang, Z., S. Schwartz, L. Wagner, and W. Miller. (2000) "A Greedy Algorithm for Aligning DNA Sequences." Journal of Computational Biology: A Journal of Computational Molecular Cell Biology 7 (1-2):203–14.

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., Tang, M., Fu, R., Li, J. and Huang. Q. (2013) "Ultra-Deep Sequencing Enables High-Fidelity Recovery of Biodiversity for Bulk Arthropod Samples without PCR Amplification." GigaScience 2 (1):4.

Zhu, B., D. G. Fitzgerald, C. M. Mayer, L. G. Rudstam, and E. L. Mills. (2006) "Alteration of Ecosystem Function by Zebra Mussels in Oneida Lake: Impacts on Submerged Macrophytes." Ecosystems 9 (6). Springer-Verlag:1017–28.

Zhu, B. (2006) "Degradation of Plasmid and Plant DNA in Water Microcosms Monitored by Natural Transformation and Real-Time Polymerase Chain Reaction (PCR)." Water Research 40 (17):3231–38.

Zuur, A., Ieno, E.N. and Smith, G.M. (2007) Analyzing Ecological Data. Statistics for Biology and Health. Springer