

The University of Maine

DigitalCommons@UMaine

Electronic Theses and Dissertations

Fogler Library

12-2016

Environmental DNA Detection of Invasive Species

Geneva York

University of Maine, geneva.york@maine.edu

Follow this and additional works at: <https://digitalcommons.library.umaine.edu/etd>



Part of the [Marine Biology Commons](#)

Recommended Citation

York, Geneva, "Environmental DNA Detection of Invasive Species" (2016). *Electronic Theses and Dissertations*. 3480.

<https://digitalcommons.library.umaine.edu/etd/3480>

This Open-Access Thesis is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.library.technical.services@maine.edu.

ENVIRONMENTAL DNA DETECTION OF INVASIVE SPECIES

By

Geneva York

B.S. University of Maine, 2013

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

(in Marine Biology)

The Graduate School

The University of Maine

December, 2016

Advisory Committee:

Michael Kinnison, Professor, School of Biology and Ecology, Advisor

Rebecca Van Beneden, Professor, School of Marine Science

Joseph Zydlewski, Professor, Department of Wildlife, Fisheries and Conservation Biology

ENVIRONMENTAL DNA DETECTION OF INVASIVE SPECIES

Geneva York

Thesis Advisor: Dr. Michael Kinnison

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Master of Science
(In Marine Biology)
December, 2016

Invasive species cause catastrophic changes to environments they are introduced into. Early detection offers the best chance at controlling the spread and mitigating any potential damages caused by the invaders. Environmental DNA (eDNA) has emerged as a more sensitive and cost effective alternative to traditional survey approaches to detection. In this study, I designed primer-probe sets for use in quantitative PCR detection of three invasive *Centrarchid* species, Largemouth Bass (*Micropterus salmoides*), Smallmouth Bass (*Micropterus dolomieu*) and Black Crappie (*Pomoxis nigromaculatus*). I surveyed 21 water bodies in Maine during two seasons (winter and spring). I designed and validated a sampling device and protocol for through-ice water sampling for eDNA. I detected target species in all lakes where they were known to be present as well as five previously unconfirmed lakes. Through hierarchical occupancy modeling I estimated the cumulative probabilities of presence, collection and detection of eDNA at three levels of the surveys (sites, samples, qPCR replicates). Although my toolsets were effective during both seasons, spring samples contained much higher concentrations of eDNA and hierarchical occupancy models showed this season to have much higher average power to detect target species than winter. Winter is still a viable season for sampling, providing fewer contamination concerns, and with a more robust sampling protocol, would be able to provide a high level of confidence of

detection. Based on my dataset, and in order to have >95% confidence of detection at each level of the survey for simple detection of presence/absence, I recommend sampling from a minimum of four sites per lake, taking three samples per site and conducting five qPCR replicates in spring and sampling from a minimum of seven sites per lake, taking five samples per site and conducting seven qPCR replicates in winter.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Michael Kinnison, for letting me come to his lab and work on this project and for all his advice along the way. I couldn't possibly have ended up in a more perfect situation. I'd like to thank my committee members, Dr. Rebecca Van Beneden and Dr. Joseph Zydlewski, for all their support and input throughout the project.

I'd like to thank all the funding sources for this project including United States Fish and Wildlife Service State Wildlife Grants Program, Maine Department of Inland Fisheries and Wildlife and Maine Agricultural and Forest Experiment Station (USDA Hatch).

I'd like to thank the whole Kinnison lab (Kristina Cammen, Brad Erdman, Jared Homola, Lauren Turinetti, Zach Wood and Wes Wright) for being awesome and always having helpful insight when I needed it, but mostly for moral boosting burger lunches. I'd like to thank Wes Wright for all his help, which I couldn't possibly list in this page, but without it I wouldn't have gotten far. Thanks to Zach for R help and Lauren for introducing me to eDNA techniques. I'd like to thank Dr. Erik Blomberg for his help with occupancy models, without which I'd have been completely lost. Thanks to Luke Lamb-Wotton for his help in the field.

I'd like to thank Merry Gallagher for all her insight and her help in figuring out where to sample. For all the sampling field help I thank both Merry again, and Tyler Grant.

Finally, I must thank the staff of UMaine, many of whom went above and beyond to help me when I really needed it on more occasions than I can remember. In particular, I'd like to thank Trish Costello, Dave Cox, Jodie Feero, Mike Hambrock, Sue Thibodeau, and Patty Singer. You all do great work and I appreciate you.

TABLE OF CONTENTS

| | |
|--|-----------|
| ACKNOWLEDGEMENTS..... | ii |
| LIST OF TABLES..... | v |
| LIST OF FIGURES..... | vi |
| | |
| CHAPTER 1: INVASIVE <i>CENTRARCHID</i> SPECIES IN MAINE AND ENVIRONMENTAL DNA..... | 1 |
| | |
| CHAPTER 2: DEVELOPMENT OF ENVIRONMENTAL DNA DETECTION TOOLS FOR FIVE <i>CENTRARCHID</i> SPECIES IN MAINE..... | 9 |
| Introduction..... | 9 |
| Methods..... | 13 |
| <i>Primer-probe set development.....</i> | <i>13</i> |
| <i>Lab validation and qPCR protocol design.....</i> | <i>15</i> |
| <i>Winter sampling device.....</i> | <i>18</i> |
| Results..... | 20 |
| <i>Primers, probes and assay conditions.....</i> | <i>20</i> |
| <i>Winter sampling device and protocol.....</i> | <i>22</i> |
| <i>Standard curves.....</i> | <i>23</i> |
| Discussion..... | 24 |
| | |
| CHAPTER 3: DETECTION OF EXOTIC <i>CENTRARCHID</i> ENVIRONMENTAL DNA IN MAINE..... | 29 |
| Introduction..... | 29 |
| Methods..... | 34 |

| | |
|---|----|
| <i>Study area</i> | 34 |
| <i>eDNA water sampling and pre-processing</i> | 36 |
| <i>Winter sampling</i> | 37 |
| <i>Spring sampling</i> | 38 |
| <i>Summer sampling</i> | 38 |
| <i>qPCR</i> | 39 |
| <i>Controls</i> | 39 |
| <i>Sequencing</i> | 40 |
| <i>Analysis</i> | 40 |
| <i>Occupancy models</i> | 41 |
| Results | 43 |
| <i>Winter sampling</i> | 44 |
| <i>Spring sampling</i> | 44 |
| <i>Summer sampling</i> | 45 |
| <i>Standard curves and starting copy number</i> | 45 |
| <i>Sequencing</i> | 48 |
| <i>Controls</i> | 49 |
| <i>Occupancy models</i> | 50 |
| Discussion | 53 |
| REFERENCES | 64 |
| APPENDICES | 69 |
| Appendix A: Primer-probe sequence alignments | 69 |
| Appendix B: Winter sampler effect | 70 |
| Appendix C: R code for hierarchical occupancy models | 71 |
| BIOGRAPHY OF AUTHOR | 74 |

LIST OF TABLES

| | |
|---|----|
| Table 2.1. Primer-Probe sets..... | 20 |
| Table 2.2. Annealing temperature optimization..... | 21 |
| Table 2.3. qPCR assay optimization..... | 21 |
| Table 2.4. Field validation of primer-probe sets..... | 22 |
| Table 3.1. <i>Centrarchid</i> eDNA sampling locations by lake, species and report category..... | 35 |
| Table 3.2. Exotic <i>Centrarchid</i> eDNA survey detections summarized by number of positive sampling sites..... | 43 |
| Table 3.3. Winter top and bottom ten estimated starting copy number/ 1 L sample..... | 45 |
| Table 3.4. Spring top and bottom ten estimated starting copy number/ 1 L sample..... | 46 |
| Table 3.5. Sequencing results from unconfirmed sites..... | 49 |
| Table 3.6. Winter occupancy model probability estimates..... | 50 |
| Table 3.7. Spring occupancy model probability estimates..... | 50 |
| Table A.1. Alignments of target species primer-probes vs. related species..... | 69 |

LIST OF FIGURES

| | |
|---|----|
| Figure 1.1. Known distribution of Largemouth Bass in Maine..... | 4 |
| Figure 1.2. Known distribution of Smallmouth Bass in Maine..... | 6 |
| Figure 1.3. Known distribution of Black Crappie in Maine..... | 8 |
| Figure 2.1. <i>Centrarchid</i> tissue collection locations..... | 14 |
| Figure 2.2. Winter sampling device..... | 19 |
| Figure 2.3. Standard curves..... | 23 |
| Figure 3.1. <i>Centrarchid</i> eDNA survey sampling locations by category..... | 36 |
| Figure 3.2. Estimated starting DNA copy number / 1 L sample by lake and season..... | 48 |
| Figure 3.3. Estimated cumulative probability of target eDNA presence at sites in a lake..... | 51 |
| Figure 3.4. Estimated cumulative probability of collecting target eDNA in samples..... | 52 |
| Figure 3.5. Estimated cumulative probability of detecting target eDNA in qPCR replicates..... | 52 |
| Figure B.1. C_q values by sampler use in spring..... | 70 |
| Figure B.2. Positive amplifications by sampler use..... | 70 |

CHAPTER 1

INVASIVE *CENTRARCHID* SPECIES IN MAINE AND ENVIRONMENTAL DNA

Early detection offers the best chance of eradicating or suppressing invasive species, of implementing timely measures to mitigate their impacts on indigenous species, and of controlling their spread (Lodge et al., 2006; Vander Zanden et al., 2010). However, detection of invasive species is often very costly and time consuming, limiting the number of sites that can be regularly surveyed to detect invaders. This is a particularly large problem for detection of many aquatic invasive species, where small numbers of individuals of an invasive species are easily concealed by aquatic environment or the challenges of sampling in water. Most state and federal agencies charged with documenting and managing fish invasions rely heavily on anecdotal reports from anglers, or other members of the public, that can be difficult to confirm, even with extensive netting, angling or electrofishing. Such sampling can also present risks to species of conservation concern. Hence, there is great need for early detection tools for invasive species that can be more widely and cost-effectively deployed than traditional approaches, and with less effects on non-target species.

Environmental DNA (eDNA) is DNA present within the environment, such as in soil or water. Its primary sources from animals are epidermal cells, feces and urine. This DNA is often present at very low concentrations within water, and is often degraded. Researchers aiming to detect eDNA have employed several different DNA amplification technologies, including standard two-primer PCR, quantitative PCR (qPCR) and next generation sequencing, however, qPCR provides the greatest specificity and sensitivity with low concentration DNA (Shokralla et al., 2012). Using species-specific primers many aquatic taxa have been detected using environmental DNA, including invertebrates, fish, amphibians, reptiles, birds and mammals (Ficetola et al., 2008; Jerde et al., 2011; Thomsen et al., 2012a; Thomsen et al., 2012b; Wilcox et al., 2013; Hunter et al., 2015). However, the specificity and sensitivity of this tool generally depend upon custom PCR approaches for any given species and such custom assays are not yet widely available for most taxa. Likewise, a sensitive and specific lab assay represents only

part of an integrated eDNA detection system, the other part being a statistically robust field and lab design with adequate replication to account for spatial heterogeneity in fish and eDNA concentrations.

Of particular concern in Maine are introductions and range expansions of multiple species of the *Centrarchidae*, or ‘sunfish’ family. In this study, I developed qPCR primer-probe sets for use in detecting eDNA of three non-native *Centrarchid* species, Largemouth Bass (*Micropterus salmoides*), Smallmouth Bass (*Micropterus dolomieu*) and Black Crappie (*Pomoxis nigromaculatus*), all three are considered invasive in Maine. All three species pose serious threats to Maine’s native fish and understanding their seasonal behavior and habitat preferences guided my further development and assessment of a field survey design to assess angler reports of illegal introductions of these species. I also developed and field tested a joint primer-probe set for two widely-distributed sunfishes native to Maine, the Redbreast Sunfish (*Lepomis auritus*) and the Pumpkinseed Sunfish (*Lepomis gibbosus*), to serve as a positive eDNA control to confirm the presence and amplification of eDNA in sites where eDNA might not detect the target exotic species. Chapter 2 of this thesis considers the development of the actual qPCR assays and validation of their specificity and sensitivity to detect my target species in lab and field samples. In that chapter I also describe the development of a novel technique to sample water for eDNA detection under the winter ice conditions that often persist for more than a third of the year in Maine. Following that initial chapter on eDNA tools, my subsequent chapter describes a proof of concept deployment of my assays for both winter and spring lake surveys to validate 21 confirmed or unconfirmed angler reports of illegal *Centrarchid* introductions in Maine. I further analyzed the data from those surveys with hierarchical occupancy models to assess the detection power of my surveys and to inform future survey strategies.

Because both of my core data chapters are written largely in the format of either a molecular methods or survey report, they focus foremost on the molecular methods, sampling approaches and survey comparisons. The information in those chapters should be understandable to a reader with some basic understanding of PCR approaches and statistics. However, this focus on the molecular and survey approaches comes at a tradeoff in the form of less details of the specific introduction histories and biology

of each of my target species. Hence, to provide more context for readers less familiar with these species in Maine, I provide some additional background here.

Largemouth Bass is a predatory fish native to the central and southeastern US. They are voracious fish predators, switching to piscivory as early as their first summer (Post, 2003). Largemouth Bass are one of the most popular game fishes in the United State and have been introduced to waters well outside their native range causing major damages in US states where they are not native as well the collapse of native fish populations in Japan, Guatemala and Africa (Zaret & Paine, 1973; ; Yonekura et al., 2004; Ellender et al., 2011). In environments with ample woody cover such as logs and fallen tree limbs, Largemouth Bass have particularly small ranges, employing sit and wait hunting strategies; in more open habitat they range further, actively seeking prey (Mesing et al.,1986; Annett, 1998; Ahrenstorff et al., 2009; Huchzermeyer et al., 2013). Their location and activity is related to season and their preference is for warmer waters than many salmonids (Hanson et al, 2007). In most Maine waters Largemouth Bass range near shore through much of the spring, summer and fall, but larger and older individuals will forage offshore and the majority of individuals retreat to deeper waters in winter. Spawning tends to take place in relatively shallow areas protected from wind and waves, with the resulting fry and young of year schooling extensively for several months before adopting a more solitary habit.

Largemouth Bass were first introduced to Maine waters more than 100 years ago, and have since become well established in many warm water lakes and ponds in southern and central regions of the state (Figure 1.1). They have become an immensely popular sportfish for local and visiting anglers, and Maine's Department of Inland Fisheries and Wildlife (MDIFW) now manages the catch with size and bag limits to protect the populations, particularly in waters deemed marginal for salmonids (Jordan, 2001). However, the popularity of the species has fostered a crisis of unplanned and illegal introductions into waterways managed for other game species. Largemouth Bass are confirmed in over 372 bodies of water, mostly in the southern and central portions of the state but reaching as far north as central Penobscot county. Given the devastating community and ecosystem effects of this species (Zaret & Paine, 1973;

Yonekura et al., 2004; Ellender et al., 2011) including impacts on other fisheries in Maine, the state seeks to monitor and halt this uncontrolled spread (Jordan, 2001).

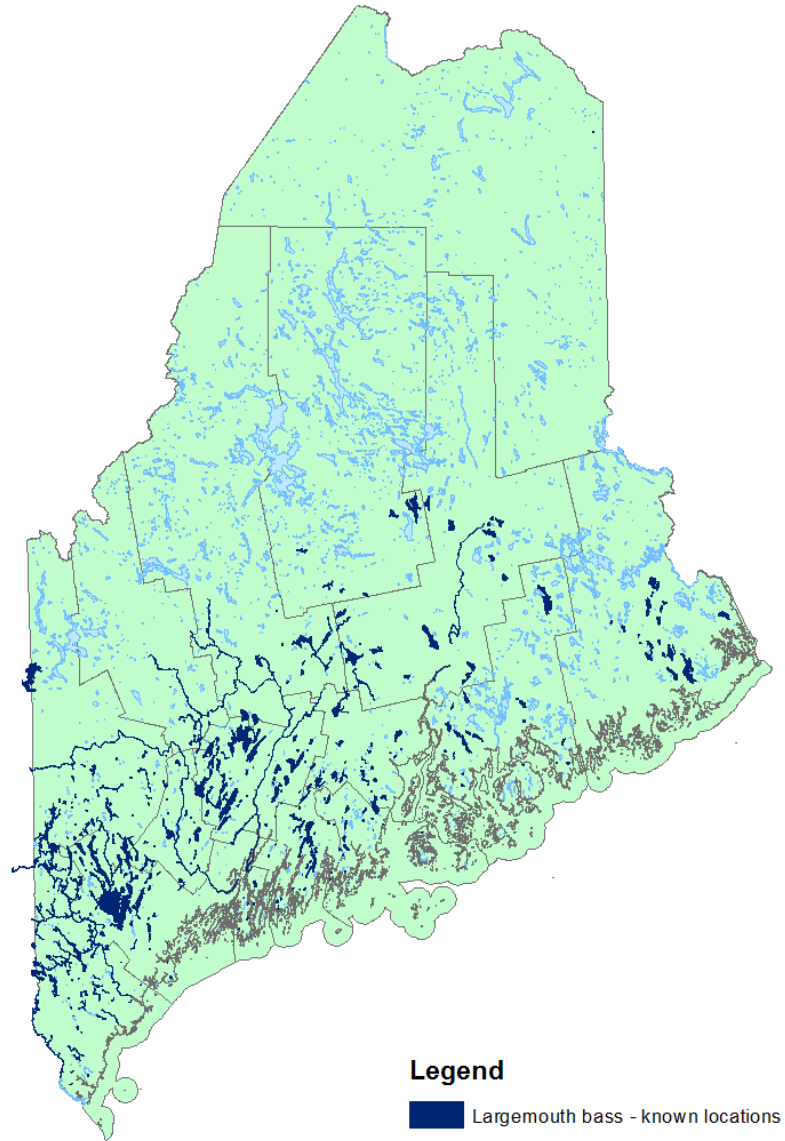


Figure 1.1: Known distribution of Largemouth Bass in Maine.

The Smallmouth Bass is a piscivorous fish native to the central and southeastern US, sharing much of its range with the Largemouth Bass but extending further north. Although the bass species share some habitat similarities, the Smallmouth Bass prefers waters with somewhat cooler average temperatures between 13-26⁰ C (Lasenby & Kerr, 2000; Vander Zanden et al, 2004) and are often associated with coarse bottomed or rocky habitats. Like Largemouth, Smallmouth Bass may be found in lakes and ponds, but they are also common in flowing rivers and streams (Dauwalter et al., 2015). Smallmouth Bass follow a similar pattern of seasonal movements to Largemouth Bass, but spend less of their year in very shallow habitats. Smallmouth often live in intermediate depth waters, but may also be found near shore, particularly where Largemouth are absent (Dewaulter et al., 2007; Brewer et al., 2011). Smallmouth Bass maintain home ranges as small as 3000 m² in productive ponds, but occupy larger ranges in less productive rivers (Minns, 1995). They are strongly connected to home territories, returning to them after capture and release kilometers away (Ridgway & Shuter, 1996).

Smallmouth Bass were introduced to Maine at the same time as Largemouth Bass and their early range expansion was actively promoted by the state. They are now wide spread throughout the state of Maine, inhabiting at least 471 water bodies, particularly within the southern and eastern regions of the state, but are confirmed in waterbodies as far north as Aroostook county (Figure 1.2). Roughly 231 lakes and ponds are co-inhabited by the two bass species, however data suggest that each species also exists as the sole bass within its habitat in many locations (Smallmouth = 240; Largemouth = 141). There is anecdotal belief that Smallmouth are often outcompeted by Largemouth in Maine, with concerns that some valued Smallmouth fisheries might be lost. Due to their popularity as sportfish, Smallmouth Bass are managed under similar catch regulations as Largemouth and, like Largemouth, their range in Maine has expanded in recent decades due to illegal introductions (Jordan, 2001).

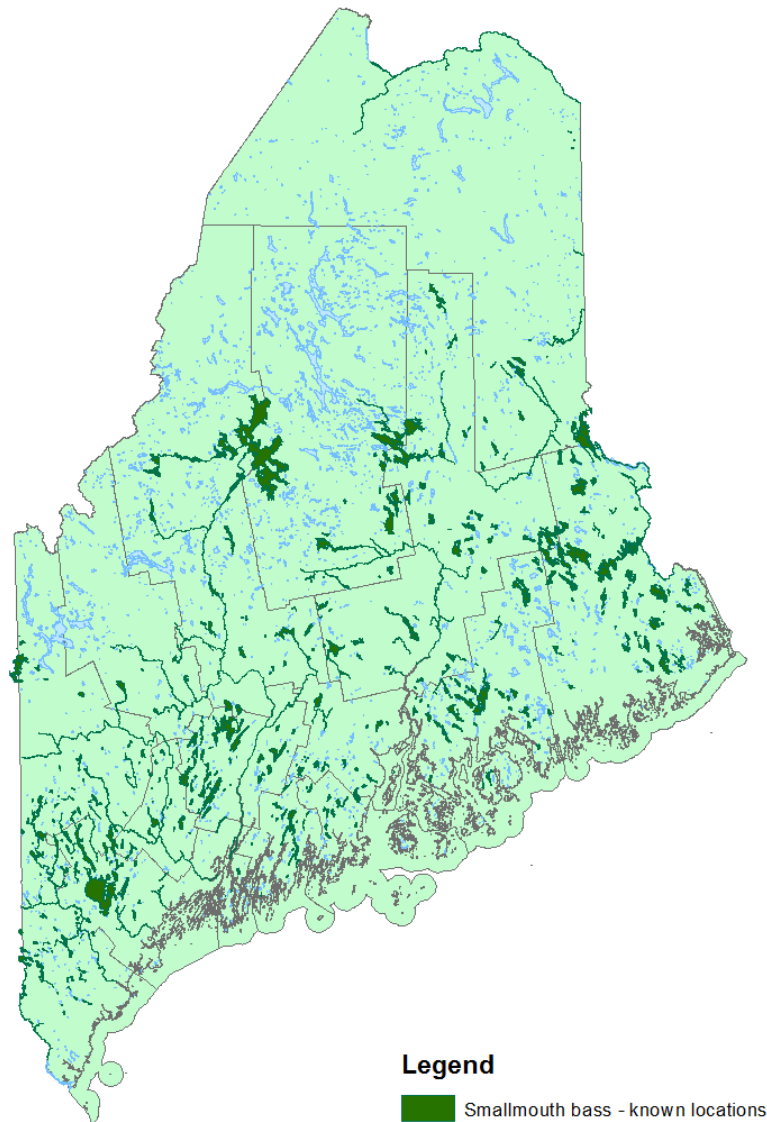


Figure 1.2: Known distribution of Smallmouth Bass in Maine.

Black Crappie are an extremely popular ‘pan fish’ native to the central US, with a growing following among Maine anglers. Crappie often co-occur in systems with Largemouth Bass, preferring lakes and ponds with still and warm waters, where they can be found along weedy shorelines (Pope & Willis, 1997; Phelps et al., 2009). They are prolific breeders, grow quickly and mature at a young age (Arslan et al., 2004). They live in habitats with variable dissolved oxygen and temperatures within their native range (Phelps et al., 2009); this adaptability and high reproductive rate make them highly capable

colonists of new systems. Seasonally, Black Crappie move into shallows to spawn in the spring, but spend much of the rest of the year at moderate depths along deep weed lines and over structure in open water. Black Crappie choose sheltered shoreline locations for nesting, preferring areas with overhanging branches rather than vegetation for cover during spawning (Reed & Perriera, 2009). Black Crappie were first stocked into Virginia Lake, in southwestern Maine in 1921. They have spread extensively through subsequent illegal introductions and colonization through river networks. It is largely managed as a nuisance species, that can compete with native fishes, including the traditionally harvested white perch (*Morone americana*), and prey on the fry of native fish species. Black Crappie were introduced in 1921, but have already spread to at least 64 lakes and ponds (Figure 1.3); (Lucas, 2002). Given the threat these species pose to native fish, Maine's Department of Inland Fisheries and Wildlife is working to keep them from moving into new waters.

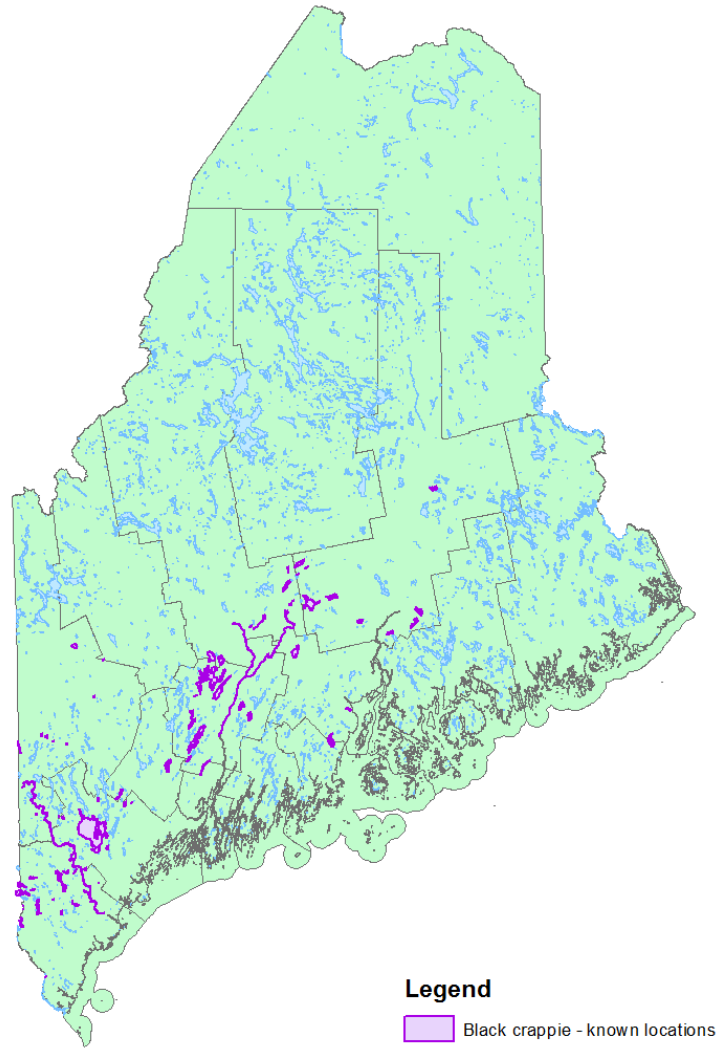


Figure 1.3: Known distribution of Black Crappie in Maine.

CHAPTER 2
DEVELOPMENT OF ENVIRONMENTAL DNA DETECTION TOOLS FOR FIVE
***CENTRARCHID* SPECIES IN MAINE**

Introduction

Detection of exotic species is often very costly and time consuming, and is a particularly large problem in aquatic environments, where small numbers of individuals are easily concealed by the limitation of direct observations in water. Traditional detection methods often begin with anecdotal reports from the public, followed by some form of capture such as angling, electrofishing or netting. These techniques are time consuming, expensive and may not be reliable when faced with new invaders at low abundance or juvenile fish (Magnusson et al., 1994; Gu & Swihart, 2004).

Largemouth Bass (*Micropterus salmoides*) and Smallmouth Bass (*Micropterus dolomieu*) were introduced to Maine waters more than 100 years ago. In that time, they have spread, or been introduced, to more than 372 and 471 waters statewide, respectively (Jordan, 2001). Both species are voracious predators and have been introduced across the globe where they've often caused immense damage to native ecosystems (Zaret & Paine, 1973; Yonekura et al., 2004; Ellender et al., 2011). In Maine, there is concern that these predators could harm sensitive native salmonid populations, including economically valuable Brook Trout (*Salvelinus fontinalis*) fisheries and federally endangered Atlantic Salmon (*Salmo salar*) in freshwater streams and rivers. Black Crappie (*Pomoxis nigromaculatus*) were introduced illegally to Maine in 1921 and has since spread to over 64 waters (Lucas, 2002). Although smaller than the basses, Black Crappie still pose a substantial threat to Maine fish by competing with native species and feeding on native fry. Alarmingly, despite strict state laws against live transport and introductions of these species, Maine continues to experience rampant illegal introductions of these species at a rate of roughly 25 new documented populations per decade (Jordan, 2001). Given limited resources, the primary agency responsible for monitoring and mitigating the spread of these exotic species, the Maine Department of Inland Fisheries and Wildlife (MDIFW), currently relies heavily on angler reports of these

species that are often poorly documented, difficult to confirm without extensive field time, or occur only after the species is well established. A cost-effective eDNA detection system could thus permit more comprehensive monitoring or quick and accurate confirmation of angler reports.

Recently, the use of environmental DNA (eDNA) for species detection has emerged as a more sensitive and inexpensive alternative to traditional sampling approaches. However, eDNA detection tools are not yet widely available and vetted for most species. In this chapter I describe the design and lab validation of quantitative PCR (qPCR) primer-probe sets for eDNA detection of three *Centrarchid* fish species that are exotic invasive in Maine, as well as for two widespread native *Centrarchid* species that can serve as positive detection controls.

Environmental DNA detection has employed several different DNA amplification technologies, including standard two-primer PCR, quantitative PCR (qPCR) and next generation sequencing. The first two of these provide the greatest sensitivities for detecting rare species, while the last affords the capacity to survey more taxa simultaneously but with less sensitivity and more assignment error (Shokralla et al., 2012). qPCR requires more initial cost due to addition of a specialized fluorescent probe, but the addition of this probe eliminates additional imaging steps of standard PCR and increases assay specificity (Piggot, 2016). In addition, qPCR provides a quantifiable signal at very low DNA concentrations that can be used to estimate the concentration of eDNA in field samples (Wilcox et al., 2013). Mitochondrial genes are generally targeted for markers in eDNA detection due to their ample among-species sequence variation and greater copy number than nuclear DNA. High copy number increases the effective eDNA signal coming from the same amount of cellular material sloughed into surrounding waters, which can be important when seeking to detect new populations at low abundances (Mills et al., 2000). To increase specificity in assays, intra-specific base pair mismatches are targeted for the 3' end of the primers (Kwok et al., 1990; Stadhouders et al., 2010). However target sequences must be chosen carefully so that they provide sufficient specificity without sacrificing PCR efficiency.

Beyond these basic lab methodology considerations, any given eDNA detection system must also ultimately be designed to facilitate detection in environmental samples, which comes with an additional

set of considerations. Like any other field detection system, eDNA monitoring systems depend on a valid field sampling scheme that is specific to the particular detection context, including the location, timing and extent of sampling. I consider many of these factors in the next chapter, but one is particularly relevant to the initial develop of eDNA tools, the specific process for collecting and successfully amplifying eDNA from a water sample. How samples are collected has a potentially large influence on the accuracy of eDNA detection. False positive eDNA detections can arise from actual presence of the target species eDNA due to unaccounted transport mechanisms, such as movements between waterways by predatory birds or dumping of ballast, bilge or live-well water from boats that travel between systems (Willerslev et al., 2003; Thomsen et al., 2012a), as well as by accidental contamination of sampling or laboratory equipment. These risks are managed in eDNA surveys by carefully planning when and where eDNA is collected to avoid transported eDNA and via development of appropriate controls to limit contamination or at least signal when and where samples are affected. False negatives, meaning cases where a species is present but its eDNA is not detected, are also a concern for a tool like eDNA where detection is dependent on successful collection, preservation, extraction and amplification of low concentration DNA. Positive controls can provide some protection against these types of process failures. One such positive control might be successful amplification a common native species known to be present in all test sites.

Maine's cold winters afford a constraint and potential opportunity for eDNA surveying that has not been explored in other eDNA studies. Essentially all currently published eDNA projects sample during the spring and summer when open water is easily accessible and presumably fish are more active (Shuter et al., 2012). However, Maine's lakes experience ice cover for as much as a third of the year, limiting the potential to survey lakes and ponds with the same methods used in other seasons. At the same time, sampling during winter ice season may provide several potential benefits, including slower breakdown of eDNA, foot or vehicle access to sampling sites, and potentially lower risk of DNA contamination from outside sources, such as boats or bird feces. Indeed, even river inflows are generally lower in winter (Kirillin et al., 2012). Currently, there is no established technique for through-ice

sampling. Additional factors to consider in winter versus open-water sampling, are the potential that the fish and their eDNA are present in different locations between seasons, the temperature challenges for field sampling gear, and the added contamination risk that might come from gear used to cut ice.

Given the needs of the State of Maine to monitor introductions or spread of exotic *Centrarchids*, and the potential for eDNA detection to provide a sensitive and cost effective approach to assist such monitoring, I developed eDNA primer-probe tools for a suite of *Centrarchids* as well as a novel field sampling system for winter eDNA surveys. Hence, the specific objectives of this study were to:

- 1) Engineer species-specific qPCR primer-probe sets for Largemouth Bass, Smallmouth Bass and Black Crappie, using mitochondrial gene sequence data for target and non-target species.
- 2) Engineer a species-pair-specific qPCR primer-probe set for Maine's native sunfishes (Redbreast and Pumpkinseed) to serve as a potential positive field eDNA control during surveys for the target exotics.
- 3) Optimize the qPCR conditions for amplifying this suite of qPCR tools and verify their efficiency and specificity.
- 4) Demonstrate the capacity of these lab tools to amplify target species eDNA in real-world water samples from sites with known species presence.
- 5) Develop a winter water sampling system to enable eDNA monitoring of the target species during a period when conventional eDNA sampling methods may be ineffective.

Methods

Primer-probe set development

For this study, I focused on mitochondrial DNA sequence variation for designing species-specific qPCR primers and probes. Using publicly available sequence data (Genbank, www.ncbi.nlm.nih.gov), I used the browser-based software Benchling (Benchling, benchling.com) to align sequence data for multiple candidate mitochondrial gene regions of my candidate species against one another and against other related species in Maine, including Green Sunfish (*Lepomis cyanellus*), Bluegill (*Lepomis macrochirus*) and Rock Bass (*Ambloplites rupestris*). From these initial alignments I selected two commonly employed target genes for subsequent primer-probe design, cytochrome b (cytb) (Minamoto et al., 2012; Takahara et al., 2013) and cytochrome c oxidase subunit 1 (co1) (Thomsen et al., 2012a). The cytb gene was chosen for initial qPCR development, with the intent that co1 could serve as a backup in the event that an effective cytb primer-probe set could not be generated. The choice of these two gene regions was based on initial evidence that they contained sufficient sequence variation between target species and other related species in Maine to be used for species-specific (or species-pair) eDNA detection.

Although mitochondrial sequence data were publicly available for my study species, most of these data were for populations well outside of Maine or New England, and one of the mitochondrial gene regions I selected (cytb) showed evidence of intra-species variability in the Genbank sequences across regions. Intra-specific sequence variability is a particular concern for my target species given their large native geographic ranges and because local sequence differences could in principle reduce the sensitivity of eDNA primer-probe sets designed using sequence variation from other regions. Hence, to ensure the greatest amount of regional sensitivity of my eDNA primer-probe sets, I collected a minimum of 3 individuals of each target species from southern and central Maine (Figure 2.1), and harvested muscle tissue samples for sequencing. I used universal fish primers from Sevilla (2007) for sequencing local tissues. Local sample sequence data were again aligned for the target species and compared with one

another and sequences of related non-target species in Maine. Sequence alignments are provided in Appendix A.

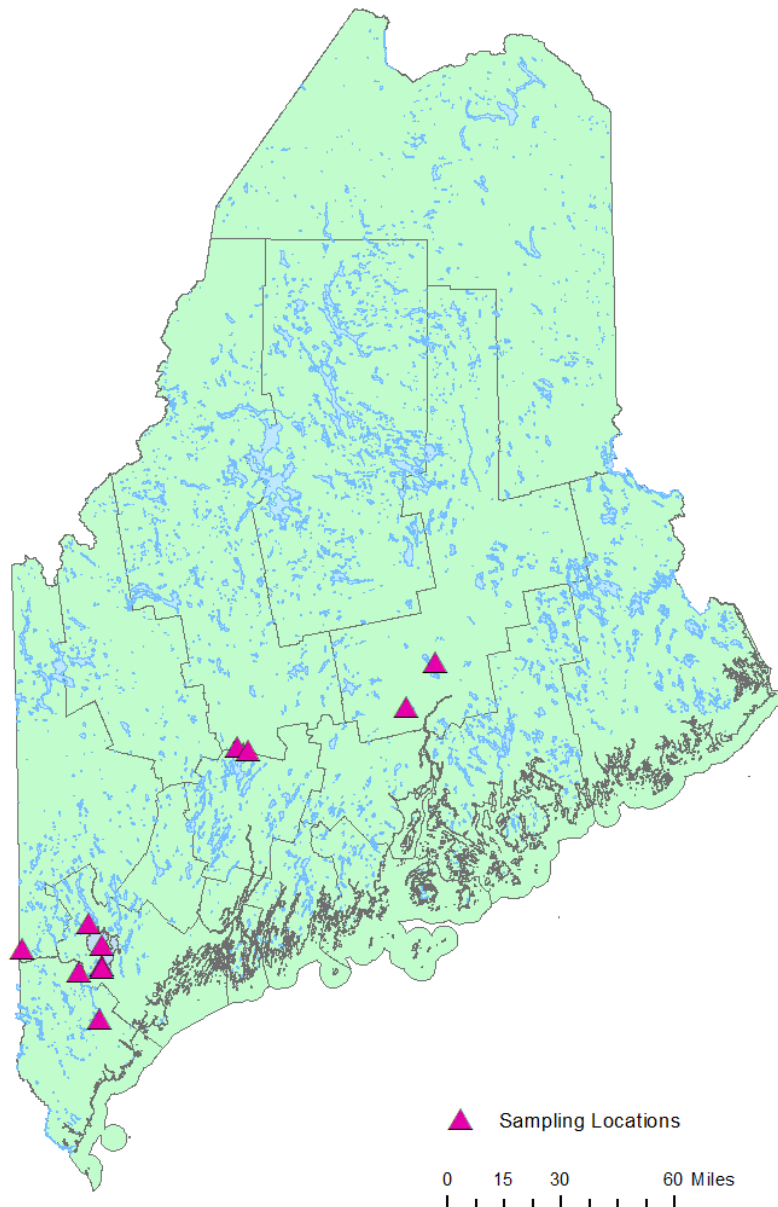


Figure 2.1: *Centrarchid* tissue collection locations

I designed eDNA primer-probe sets using Benchling's primer and probe development tool (Benchling, www.benchling.com). Each eDNA primer-probe set included forward and reverse primers,

creating amplicons of between 95-150 bases in length, with a Taqman minor groove binding (MGB) probe seated in the sequence between the primers. The Taqman MGB probe has a moiety on the 3' end which folds into the minor groove of the DNA duplex. This creates a strong and stable hybridization, and a high melting temperature. MGB probes can thus be designed with shorter sequences, creating greater specificity in the probe (Kutyavin et al., 2000). In addition, probes were designed with a non-fluorescent quencher (NFQ). Taqman MGB-NFQ probes have extremely low background fluorescence, of particular importance when amplifying environmental samples with low DNA concentrations (Kutyavin et al., 2000; Josefsen et al., 2009). To maximize species-specificity, intra-species mismatches were targeted for the 3' end of the probe, keeping in mind Taq polymerase's lack of 3' to 5' exonuclease activity (Kwok et al., 1990; Stadhouders et al., 2010). I cross-checked all primers against the NCBI database for all organisms using PrimerBLAST.

Lab validation and qPCR protocol design

Using DNA extracted from known tissue, I first tested the amplification and specificity of my designed forward and reverse primer sets against all five target species, plus one newly reported non-native *Centrarchid* (Bluegill, *Lepomis macrochirus*), using standard PCR and ethidium bromide gel staining for visualization of presence or absence of an amplification fragment. These initial PCRs were conducted at 56^o C annealing temperature, based on the predicted thermal melting points of the primers (addition of a probe raises the annealing temperature for qPCR). Once I was sure all primer sets effectively amplified their target species, and only their target species, via standard PCR, I proceeded to optimize amplification conditions for qPCR. All qPCR reactions were amplified and quantified using a Bio-Rad CFX96 Real-Time System thermal cycler on 96-well qPCR plates. Reactions were run in 20 µl volumes using a fixed 10 µl of Taqman Environmental Master Mix 2.0 (Applied Biosystems) per reaction.

Use of lower annealing temperatures for primers and probes often favors more sensitive amplification of low concentration DNA but comes with a tradeoff of lower specificity of the DNA sequence that is amplified. Likewise, primer concentration may affect mis-priming and primer-dimer formation. To determine optimal primer-probe concentrations and annealing temperatures, I ran a six-part, five-fold dilution series of synthetic gene fragments (31250, 6250, 1250, 250, 50, 10 copies/ μ l) (Mikeksa & Dobrovic, 2009; Taylor et al., 2010) and conducted standard curve analyses to estimate qPCR efficiencies. First, I ran all primer-probe sets at the same concentration (10 μ M / 5 μ M) along a thermal gradient between 56-63^o C with 4 technical replicates, as shown in Table 2.2. During qPCR, the exonuclease activity of Taq polymerase cleaves the fluorescent tagged probe, separating the dye from quencher (Holland et al., 1991). Fluorescent signal is created each time target DNA is amplified, and it can be measured, creating a way to quantify the amount of DNA amplification that has occurred during the PCR reaction (Livak et al., 1995). As the PCR reaction runs, amplification increases as does the fluorescent signal, creating a rapidly growing fluorescence curve, until reagents are consumed below effective levels. The point in the cycle at which the total fluorescence moves beyond a baseline threshold to observable levels is the quantitation cycle or C_q. Since the fluorescence is directly linked to amplification, the more initial DNA template available, the earlier the fluorescence reaches observable levels (Heid et al., 1996). I defined a universally optimal annealing temperature for all assays as the temperature associated with the lowest average C_q per dilution still not showing cross-species amplification for any of the eDNA assays as described by Taylor (2010). I then ran all primer-probe sets at a range of concentrations (20 μ M/10 μ M, 20 μ M/ 5 μ M, 15 μ M/ 5 μ M, 10 μ M/ 5 μ M) (Table 2.3). Each primer concentration had four technical replicates per primer-probe set concentration. I determined the most efficient concentration via standard curve analysis (Mikeksa & Dobrovic, 2009). I then re-ran the most optimal primer-probe concentrations at the universal optimal temperature (60^o C), with 4 technical replicates, as an additional confirmation and quantification of overall assay efficiency. All

dilution series for standard curve analyses included six five-fold dilutions (31250, 6250, 1250, 250, 50, 10 copies/ μ l).

Once I established effective lab eDNA amplification conditions, I further tested my assays for their ability to amplify eDNA in 1L field-collected water samples from sites with known target species presence (or absence). These water samples were filtered in a DNA decontaminated lab via vacuum filtration through glass fiber filters (Whatman 1.5-micron pore, 5.5 cm diameter). Potential eDNA on the filters was extracted with a Qiagen blood and tissue kit (Qiagen, Valencia, CA) and eluted to a volume of 100 μ l in nuclease free water. All qPCR amplifications of field samples were run with Taqman Environmental Master Mix 2.0 (Applied Biosystems) to reduce the potential for inhibition in field water samples and to ensure uniformity from lab optimization to field use (Jane et al., 2015). Reactions were 20 μ l in total volume, 10 μ l of Taqman Environmental Master Mix 2.0 (Applied Biosystems), 2.625 μ l of primer-probe at 10 μ M – 5 μ M concentrations, 4.375 μ l of nuclease free water and 3 μ l of template. Primer-probe concentrations were 10 μ M - 5 μ M for all assays. The thermal cycle protocol was initial denaturation at 95⁰ C for 10 minutes followed by 50 cycles of denaturation at 95⁰ C for 10 seconds, annealing and extension at 60⁰ C for 30 seconds.

During a large field sampling survey (Chapter 3) I ran a series of six 5-fold dilutions of synthetic gene fragments corresponding to the target species DNA fragment set being tested. This dilution series was run on all survey plates and served as both a positive amplification control and as a standard calibration curve relating the C_q of sample amplification to eDNA copy number. However, the resulting 57 standard curves also provide a means to verify eDNA assay consistency and qPCR efficiency. I plotted the C_q values for each dilution against the log of the gene fragment concentration for each dilution series on each plate. Since DNA template ideally doubles each cycle during amplification, the C_q values of known concentrations of template should vary by a known amount. The slope of the linear regression of C_q versus the log of the gene fragment concentration, if the efficiency of the primer-probe set is 100%, should be -3.332. The formula for calculating the efficiency is: $-1+10^{(-1/slope)}$.

Winter sampling device

During pilot testing, I failed to detect eDNA of one target species in 1L water samples from two waterbodies where the target is known to be present. These initial tests consisted of cutting through ice to sample water at the lake surface. I determined that the probable basis for these initial detection failures is the winter behavior of our focal species and the greatly reduced mixing of lake water during the winter ice period. The *Centrarchids* I targeted primarily move to deep water habitats during winter and take up residence in patchily distributed high concentration areas. Lakes with ice cover are protected from mixing by wind and experience a pattern of reverse stratification (4°C on bottom and colder temperatures at top), that could trap target species eDNA in deeper waters. Hence, in line with prior recommendations that eDNA water sampling target probable habitats where species and their eDNA are concentrated (Eichmiller et al., 2014), I sought to employ a field sampling approach that would facilitate targeted sampling of deeper lake waters during the winter ice period. However, no published eDNA studies exist outlining methods for sampling fish eDNA through winter ice, or at depths beyond a few meters, necessitating development of a novel winter sampling tool and protocol.

From a logistics standpoint, an effective tool for winter eDNA sampling must be portable, be deployable through ice, target water at designated depths, be resistant to malfunction under potentially freezing conditions, and present low risks of eDNA sample contamination. Optimally, such a tool would also be widely available or inexpensive so as to facilitate future monitoring by agencies, NGOs or citizens groups with limited resources or technical capacity. Currently, the most common alternative to direct bottle sampling of surface waters for eDNA analysis is use of a portable peristaltic water pump and tubing to draw a known volume of subsurface water through a filter apparatus in the field. However, portable peristaltic pumps are not widely available and the combination of freezing conditions and long lengths of tubing (6 to 20 meters) would likely make such an approach mechanically unreliable. Van Dorn bottles are commonly used in limnology to sample water at depth, but these devices are expensive and not widely available. Likewise, deployment of a Van Dorn bottle in winter requires drilling a large hole in the ice (e.g., 20 cm diameter), necessitating use of a large diameter manual, gas or electric ice auger. Long

lengths of tubing, Van Dorn bottles, and large mechanical ice augers were all further deemed to present challenges for efficient DNA decontamination when sampling multiple sites. I thus sought to develop a new sampling apparatus that could be constructed from inexpensive and widely available hardware and that can be deployed through holes in ice created with tools small enough to efficiently decontaminate.

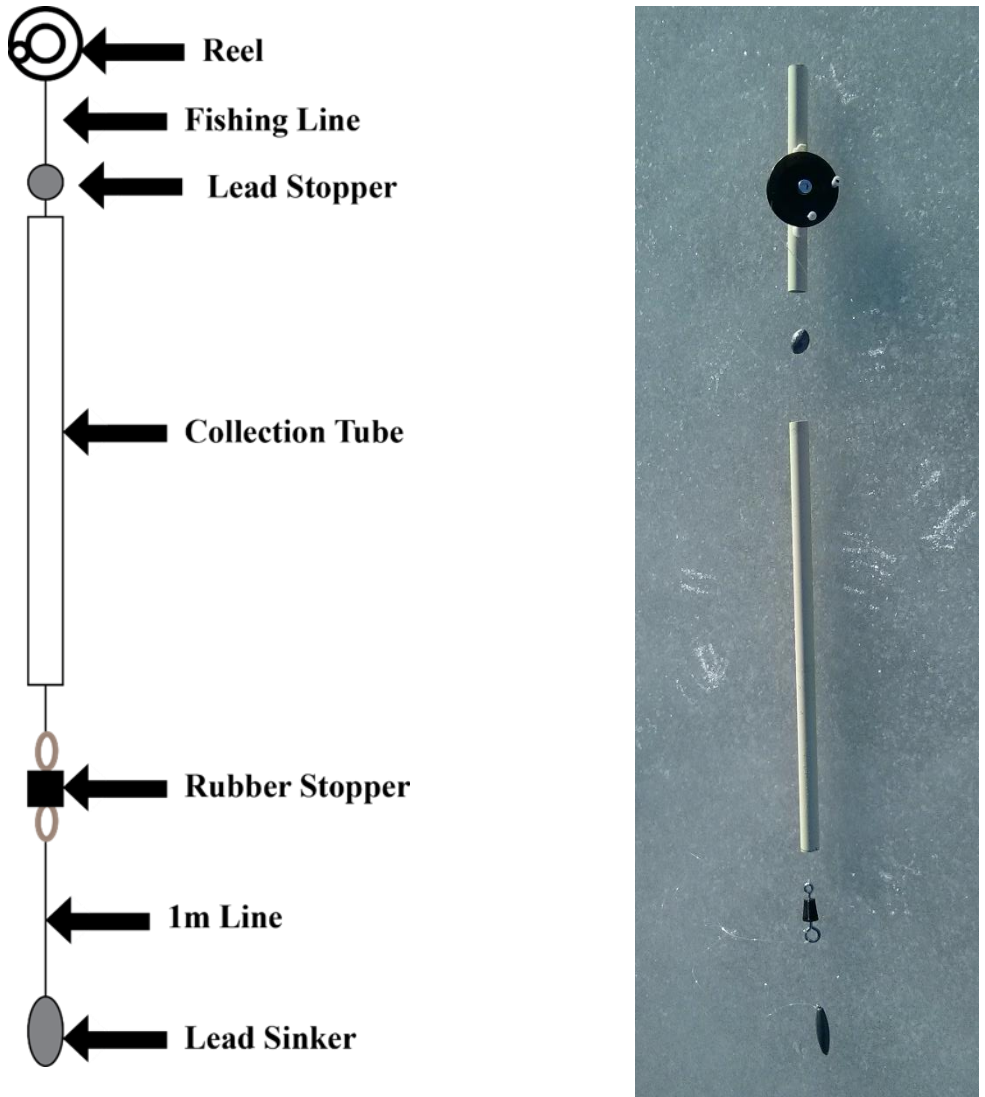


Figure 2.2: Winter sampling device

Results

Primers, probes and assay conditions

The eDNA assays for Largemouth Bass (LMB), Black Crappie (BC), and the sunfish pair (SUN) utilize the mitochondrial gene *cytb*, while the Smallmouth Bass (SMB) assay uses the mitochondrial gene *co1* (Table 2.1). Initially I planned to design all assays using the *cytb* gene, which is the most commonly used gene for eDNA marker sets, but I had difficulty designing the SMB set on the *cytb* gene with desired T_m while still ensuring enough cross-species mismatches and maximum mismatches on the 3' ends. Thus, I switched to the *co1* gene for the SMB assay, which provided ample variation among related *Centrarchids*. Subsequent lab testing validated the specificity of the final assays for all the target exotic *Centrarchids* when run under the specific qPCR conditions described below. The sunfish assay effectively amplified DNA from both native sunfish species, but not the DNA of the other *Centrarchid* species in Maine, consistent with my goal of designing a universal native sunfish assay.

Table 2.1: Primer-Probe sets. Forward and reverse primers and MGB probes for three target species, Black Crappie (BC), Largemouth Bass (LMB) and Smallmouth Bass (SMB). Redbreast Sunfish and Pumpkinseed Sunfish DNA both amplified by SUN assay.

| Assay | Forward Primer (5'-3') | MGB Probe (5'-3') | Reverse Primer (5'-3') |
|--------------|-------------------------------|------------------------------------|-------------------------------|
| BC | GCCTCTGCTTGCCACCCAAAT | GCAACTGCCTTCTCCTCCGTAGCA | CCGCAACATTCATGCCAATCGG |
| LMB | CGCTGCCGCCACAGTAATCCAT | CCCCCTGGGACTAAACTCTGACGC CG | TCGCAGCTCTCCTCATTGCCCT |
| SMB | GGGTGTCTCTCCATCCTAGGGGCC | TCAGACACCCCTGTTTGTGGTCC GTCTT | TTATCGCTCCCAGTCTCGCTGC |
| SUN | CTCTACAAGGACCTCCTCGGCTTT | GCCCTCCTATTGCCCTGACTTCCC TGGCAC | TGCTCGGGGACCCAGACAACCT |

The universal annealing temperature identified for all assays was 60⁰ C, and the overall thermal cycle protocol for further lab and field validation tests was: initial denaturation at 95⁰ C for 10 minutes followed by 50 cycles of denaturation at 95⁰ C for 10 seconds, annealing and extension at 60⁰ C for 30 seconds, following the Taqman Environmental Master Mix 2.0 suggested protocol and optimal annealing temperature determined through testing (Table 2.2). I considered a positive amplification a curve beginning before cycle 45 (Wilcox et al., 2013). Although there was some evidence of variability among assays in optimal primer-probe concentrations, I found that all assays were at or near optimum efficiency with primer-probe concentrations of 10 μ M and 5 μ M, respectively. The one exception to this were the

sunfishes, for which optimal primer-probe concentrations were somewhat higher (15 μ M and 5 μ M). However, given that the sunfish assay was developed for use as a positive field eDNA control for two common, and frequently abundant, species I opted to set the primer-probe concentrations of all species at the same 10 μ M and 5 μ M level, recognizing that doing so may make the sunfish positive control assay somewhat conservative.

Table 2.2: Annealing temperature optimization: Lowest C_q and average C_q from highest gene fragment dilution, Cross species amplification considered positive at any dilution. Ideal temperature would be that with lowest C_q with no cross-species amplification.

| Assay | Temp of lowest C_q / C_q | Temp with species cross amplification | Average C_q at 60 |
|-------|------------------------------|---------------------------------------|---------------------|
| BC | 57/26.12 | <56/no amp | 26.81 |
| LMB | 56/22.54 | 58.4 | 23.13 |
| SMB | 56/25.28 | 56 | 26.12 |
| SUN | 56/21.11 | 56 | 21.70 |

Table 2.3: qPCR assay optimization: Slope of linear regression of C_q versus log of gene fragment dilution, qPCR efficiency calculated via : $-1+10^{(-1/slope)}$. Ideal primer-probe concentration selected as efficiency closest to 100% with r^2 of 0.99.

| Assay | [Primer μ M /Probe μ M] | slope | Efficiency % | r^2 |
|-------|---------------------------------|--------|--------------|-------|
| BC | 20/10 | -3.578 | 90.3 | .98 |
| BC | 20/5 | -3.268 | 102.6 | .99 |
| BC | 15/5 | -3.448 | 94.9 | .99 |
| BC | 10/5 | -3.254 | 102.9 | .99 |
| LMB | 20/10 | -3.539 | 91.7 | .99 |
| LMB | 20/5 | -3.448 | 94.9 | .99 |
| LMB | 15/5 | -3.257 | 102.8 | .93 |
| LMB | 10/5 | -3.318 | 100.2 | .99 |
| SMB | 20/10 | -3.597 | 89.7 | .96 |
| SMB | 20/5 | -3.176 | 106.5 | .99 |
| SMB | 15/5 | -3.447 | 95.0 | .98 |
| SMB | 10/5 | -3.278 | 101.6 | .99 |
| SUN | 20/10 | -3.209 | 104.8 | .99 |
| SUN | 20/5 | -3.390 | 97.1 | .99 |
| SUN | 15/5 | -3.249 | 102.9 | .99 |
| SUN | 10/5 | -3.257 | 98.4 | .99 |

I tested assay sensitivity by running six 5-fold serial dilutions of synthetic gene fragments (31250, 6250, 1250, 250, 50, 10 copies/ul). With this information, I ran standard curve analysis (Table 2.3), establishing ideal annealing temperature across species. Utilizing the synthetic gene fragments, I

determined that the assays are all able to detect down to 10 copies of DNA per reaction (1/2 copy/ μ l reaction). I did not test lower concentrations of DNA due to accuracy of the pipettes used for dilutions.

Table 2.4: Field validation of primer-probe sets: Water samples collected from sites of known target presence. Expected results/actual result. +=positive result, -= negative result, NT=not tested.

| Assay | Penobscot River | Pushaw Lake | Hermon Pond | Mud Pond | Floods Pond |
|--------------|------------------------|--------------------|--------------------|-----------------|--------------------|
| BC | NT | -/- | +/+ | -/- | -/- |
| LMB | NT | +/+ | +/+ | +/+ | -/- |
| SMB | +/+ | +/+ | +/- | +/- | -/- |
| SUN | +/+ | +/+ | +/+ | +/+ | +/+ |

Winter sampling device and protocol

In order to access the bottom water through a 2.54 cm diameter hole in the ice, I constructed a simple whole-water collection device (Figure 2.2) from PVC pipe, a conical rubber stopper and fishing gear. The device consisted, from bottom to top, of a 57-gram lead torpedo sinker, a 2.2 cm diameter rubber stopper equipped with top and bottom steel eyelets, a 2.2 cm diameter PVC pipe which was .45 m in length, a 57 gram lead egg sinker, and 50m of 13.5 kg fishing line. Between the bottom sinker and the rubber stopper I strung 1m of fishing line. Remaining fishing line was attached to the top eyelet on the rubber stopper and threaded through the PVC pipe and egg sinker. The end of the line with the torpedo sinker and stopper was then lowered down the ice hole until the sinker contacted the lake bottom at which time the PVC pipe was deployed down the hole on a taught line. Once the PVC pipe was detected to seat itself on the stopper, by the sensation of impact on the line, the egg sinker was then dropped down the line to both seal the top of the tube and further seat the PVC pipe on the stopper. Deployed in this fashion, the water in the sampler effectively represented a coring of the water column from approximately 1.0 to 1.5 m from the lake bottom. The sample apparatus was then retrieved to the surface with the fishing line and emptied into a sterile collection bottle by lifting the top egg sinker and upending the contents. Optionally, a simple spool/reel can be added to the apparatus to assist with line retrieval and management. All parts of this sampling system, including both the water sampler and auger bits used to drill ice, are readily DNA decontaminated by immersion in a 10% sodium hypochlorite (Chlorox bleach) bath in the field or lab, although the inexpensive nature of the sampler can also facilitate disposable use.

This sampling apparatus and protocol was subsequently confirmed to effectively collect 1 L water samples containing eDNA for the target species in sites with their known presence. Moreover, while there is no *a priori* reason to expect water samples collected at a given depth with this apparatus should contain higher or lower concentrations of eDNA than water samples collected with other approaches, I conducted a paired sampling study to confirm this assumption (Appendix B).

Standard curves

Synthetic gene fragments amplified generally as expected on all plates, though 6 wells out of 342 failed to amplify. Figure 2.3 shows the 57 individual standard curves, composed of 6 points each, as well as the mean standard curve in red. The overall PCR efficiency across plates, based on the slope of the mean curve (-3.23), was estimated at 103.9%, with a 95% CI of 99.6-106.8% efficiency based on the variability among plates.

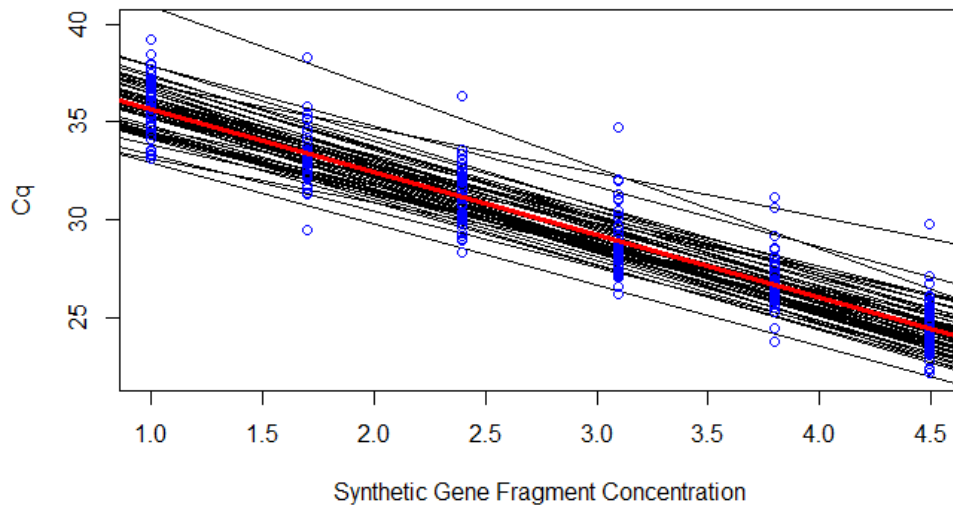


Figure 2.3: Standard curves: All 57 standard curves from field sample plates. Log of gene fragment concentration on x-axis, C_q values in y-axis. Amplification points in blue, mean standard curve in red.

Discussion

During the first phase of the project I successfully designed three qPCR primer-probe sets for single *Centrarchid* species and one set which amplifies eDNA from two native sunfish species. All sets amplify short (96-150 bp) sequences of mitochondrial DNA from one of two genes, *cytb* or *co1*. I assessed the primer-probe concentrations and thermal conditions of the PCR protocol for maximum efficiency of the target exotic species assays. When I designed the primer-probe sets I targeted the same specific melting temperature for all assays. Although some of the assays proved specific enough to be run at lower temperatures, I standardized the temperature across assays so they could be run on the same plate, leaving open the possibility for multiplexing. The temperature standardization resulted in acceptable qPCR efficiencies. I confirmed the efficacy of the assays and protocol with known tissue samples and environmental samples of known presence/absence. Finally, I designed and tested equipment and a procedure for eDNA water sampling through ice. In the rest of the discussion I will consider how my assays compare to assays available for other species and the utility of my winter sampling system for *Centrarchid* surveys in Maine.

Amplicon size affects qPCR efficiency; smaller amplicons are one of the factors that lead to more efficient qPCR reaction. Given that qPCR requires a probe to sit between the forward and reverse primers, the amplicons for qPCR assays are often longer than those for standard PCR assays. My assays produced amplicons between 96 and 150 base pairs in length. These size ranges are comparable to amplicons produced from other qPCR based eDNA assays, which are generally 140-150 base pairs in length (Wilcox et al., 2013; Eichmiller et al., 2014; Hunter et al., 2015). My assays in initial lab tests had efficiencies between 98.4% and 102.9% in the lab (Table 2.3) and 103.9% in the field, compared to other eDNA assays for other species which have published efficiencies between 86 and 99%, which include field tests (Eichmiller et al., 2013; Wilcox et al., 2013; Hunter et al., 2015; Lacoursière-Roussel et al., 2015). My assays were also able to detect DNA concentrations as low as 2 copies of DNA per reaction, which was the lowest detectable concentration reported by Wilcox (2013) and which approaches the

theoretical lower limits of detection. Hence, even prior to field implementation, the assays designed in this study hold up well against others already tested in the field.

Although one of the main benefits of using the qPCR assays is the sensitivity and ability to detect very low concentrations of DNA, this can also be one of its greater risks. Even a very small amount of contamination can lead to false positives. Field and lab controls are typically employed to detect any potential contamination, and careful selection of survey sites can also limit the likelihood of sampling eDNA introduced to the water body by outside means, rather than produced by fish present. Winter sampling could in principle be one way to reduce the chance of outside contamination given that ice limits the amount of DNA introduced by animals, boats or angling gear as well as reducing inputs off the surrounding landscape. I conducted basic validation of the ability of my assays to amplify fish eDNA in winter. These winter sampling approaches were predominantly successful, but notably they were most successful when species movements and water conditions were taken into account. Specifically, the SMB assay failed to amplify the target in two locations where I expected to find this species, Hermon Pond and Mud Pond. At the same time, my pilot field testing of my sunfish eDNA primer-probe set (and of other species assays), was successful at these same sites, suggesting that SMB eDNA would likely have been detected had it been present in samples. This not only demonstrated the value of including positive field eDNA controls to rule out sample degradation, it led me to test Penobscot River water samples that were taken in the early fall of the preceding year when there was no ice and the fish were presumably more active and water better mixed. These samples yielded excellent results, proving the efficacy of the assay, but leading to some concern over how sampling should be conducted in winter to maximize the likelihood of encountering target species' eDNA.

My initial winter samples for winter testing were taken from the surface through holes cut in the ice, but the lack of water movement, deep water migration of *Centrarchids* in winter, and reverse stratification of lakes in winter, could easily limit the amount of eDNA that is encountered for SMB and other species in surface samples (Kirrilin et al., 2012; Peat et al., 2016). Hence, I altered my sampling protocol to sample water near the bottom in 6-20 meters of water, where I believed SMB and other

Centrarchids are most likely to reside in winter (Peat et al., 2016). The improved sampling scheme proved successful in follow-up tests (Pushaw Lake, Table 2.4). Hence, it looks promising that the winter sampling tool and procedures developed here could open up the opportunity for effective winter eDNA water sampling for many other species and study systems, again provided the tool is deployed in a fashion that matches the biology of the species and hydrological processes operating during the sampling period. In Chapter 3, I actually quantify my level of success in sampling eDNA under these revised protocol conditions compared to typical surface water sampling in other seasons.

Notably, although I designed my winter sampling device with through-ice sampling in mind, the need to sometimes sample deeper waters is not unique to winter and the tool itself could also be utilized for at-depth sampling in any season. Most eDNA samples collected during open water periods are taken at the surface during periods when target species are expected to use that habitat or lake processes promote mixing, but that can be a narrow window of time for some species, such as the Lake Trout studied by Lacoursière-Roussel (2015). During the open-water season lakes frequently experience strong thermal stratification, and many species that make use of cooler hypolimnetic waters. Such species could again be difficult or impossible to detect in surface samples during those periods. Although there are other water collection devices designed for sampling at specific depths, like Van Dorn bottles, the sampling device from this study has several advantages in that it is constructed from items easily purchased at a hardware store (<\$10), can reliably sample at any desired depth, and it can be easily and fully decontaminated (or even disposed).

Although the assays developed here have shown to be very efficient and able to amplify low concentrations of DNA, and I succeeded in developing a winter water sampling system, the utility of these tools for field surveys will depend on pairing them with an effective field survey scheme. Currently, there is no such thing as a standardized sampling scheme in eDNA studies. Previous studies have taken as few as one and as many as twenty-four water samples per water body (Takahara et al., 2013; Eichmiller et al., 2014). Sample volumes are commonly 1 L, like my own, though some studies have used volumes as small as 15 ml while still being able to detect eDNA (Thomsen et al., 2012a;

Takahara et al., 2013; Eichmiller et al., 2014; Hunter et al., 2015; Lacoursière-Roussel et al., 2015). Beyond sample volumes, Hunter (2015) suggests that sample number is more important for detection of rare species than is qPCR replicate number. They used hierarchical occupancy models to determine that increasing the number of samples increases the probability of getting a positive detection more than increasing the number of qPCR replicates based on their ability to detect radio-tagged Burmese Python (*Python molurus bivittatus*). Considering the power of the assays to detect DNA concentrations as low as 2 DNA copies per reaction, and a lack of independence of qPCR replicates from a given sample extraction, many qPCR replicates may generally not be as useful as multiple samples. However, Ficetola (2015), through simulation, found that increasing PCR replicates decreases the both false positive and negative amplifications, especially at very low concentration of DNA. Eichmiller (2014) also found that eDNA of the common carp (*Cyprinus carpio*) was not homogenous throughout water bodies or the water column, but rather concentrated where fish were located. In contrast, Lacoursière-Roussel (2015) sampled soon after ice out in the early spring, when the water column is well mixed and not yet thermally stratified and found that although Lake Trout (*Salvelinus namaycush*) eDNA was more concentrated where the fish were more abundant, it was found widely throughout any given lake. These various results of prior studies emphasize that where, when and how many water samples are taken can be important to the odds of species detections, particularly with respect to where target species are likely to be located and the hydrology of systems during the sampling period. Hence, while the results I present in this chapter describe the essential ground work of tool development necessary for future eDNA surveys, these findings are not sufficient in themselves to assess the full effectiveness of such tools.

In closing, my preliminary field tests showed that all designed assays are effective and selective at detecting their target species in real water samples and did not amplify in samples where the targets were known to be absent. The winter sampling device allowed me to collect eDNA from target habitats near the bottom, when I was unable to detect it near the surface. However, this basic survey consisted of only a few samples from five field locations. As I have noted, a comprehensive eDNA sampling system requires more than just effective eDNA and water sampling tools, it requires a field sampling design with

the spatio-temporal statistical power to assure detection of target species at a specific confidence level. In the next chapter I use an expansive field survey of lakes with reported populations of my target *Centrarchids* to demonstrate the real-world efficacy of my eDNA tools for exotic species monitoring and to provide quantitative recommendations of a field survey design with a desired statistical power.

CHAPTER 3

A MULTI-LAKE AND MULTI-SEASON SURVEY OF EXOTIC *CENTRARCHID* ENVIRONMENTAL DNA IN MAINE

Introduction

Invasive fish species alter aquatic communities in many ways, such as causing changes to trophic interactions and nutrient cycling (Strayer, 2010; Strayer, 2012) and even eliminating native competing or prey species (Zaret & Paine, 1973; Pimentel et al., 2005). The freshwaters of the state of Maine are widely regarded for their comparatively pristine conditions compared to waters throughout much of the remaining eastern United States. A relatively large number of lakes and rivers support fisheries for regionally indigenous salmonids, including Brook Trout (*Salvelinus fontinalis*), Landlocked Salmon (*Salmo salar*), Lake Trout (*Salvelinus namaycush*) and Arctic Charr (*Salvelinus alpinus*) (Frost, 2001; Johnson, 2001; Boucher, 2004; Bonney, 2009). These same waterways often support other fishes that are broadly considered representative of northern ‘coldwater’ fish assemblages, such as cusk (*Lota lota*) and rainbow smelt (*Osmerus mordax*). In Maine, aquatic invaders have the potential to have strong negative effects on these sensitive and regionally significant fish communities, and the Maine Department of Inland Fisheries and Wildlife (MDIFW) considers invasive species introductions to be a top threat to the sustainability of existing fisheries (MDIFW, 2015). Of particular concern in Maine are illegal introductions and range expansions of multiple species of the *Centrarchidae*, or the ‘sunfish’ family. In this study, I conducted proof of concept eDNA surveys for three high-priority species of exotic *Centrarchids*, in two different sampling seasons (winter and spring), and used the resulting data to estimate eDNA detection probabilities and recommend robust field sampling designs for future eDNA surveys.

Unchecked spread of Largemouth Bass (*Micropterus salmoides*), Smallmouth Bass (*Micropterus dolomieu*) and Black Crappie (*Pomoxis nigromaculatus*) poses serious threats to Maine’s native fishes.

Moreover, there are concerns that spread of these species creates conflicts with important fisheries. For example, Largemouth Bass appear able to negatively impact valued Smallmouth Bass fisheries and Black Crappie may negatively compete with White Perch (*Morone americana*) and Yellow Perch (*Perca flavescens*);(Lucas, 2002). These three species have been legally or illegally introduced to new waters in Maine at a faster rate than any other gamefishes. Between 1868 and 2001 the state of Maine documented 840 new introductions of the two bass species alone. Continued high rates of illegal establishment of new populations of these species place a significant strain on MDIFW's resources and capacity for early detection.

Detection is the first critical step in managing the problems of invasive species. In Maine, as with most places, detection most often begins with anecdotal accounts from the angling public. Current approaches for confirming exotic species usually take the form of visual or capture surveys, including snorkeling, netting, angling or electrofishing. These methods often require many person hours, promote avoidance behaviors in fish, and some cause harm or death to non-target fishes of recreational or conservation value. These techniques can miss juvenile fish and can completely miss species with very low populations, such as those of recent invaders (Magnusson et al., 1994; Gu & Swihart, 2004). Because of the cost and low efficiency of these approaches, confirmation surveys themselves are often not conducted for less-credible reports. Arguably, these major limitations of invasive species reporting and confirmation could be overcome by a much cheaper, sensitive and reliable detection system. Indeed, a sufficiently inexpensive but effective tool might not only facilitate more effective confirmation of anecdotal reports of new introductions, but even permit a fundamental shift in strategy from reliance on incidental detections and confirmation, to one of widespread and consistent monitoring. Environmental DNA (eDNA) is emerging as a species detection tool that could make this transition a reality.

Environmental DNA detection has employed several different DNA amplification technologies, including standard two-primer PCR, quantitative PCR (qPCR) and next generation sequencing. qPCR is the most sensitive, able to detect DNA down to only 2 copies of template DNA in a reaction, and requires fewer steps than standard PCR due to absence of gel visualization (Shokralla et al., 2012; Wilcox et al.,

2013). Regardless of the particular amplification technology, it is important that putative eDNA detection assays must be designed and evaluated for sensitivity and specificity. I described and conducted this design and evaluation processes in the preceding chapter. However, like any other detection system, a fully integrated eDNA detection system depends on much more than just the sensitivity and specificity of the technological instrument itself. All monitoring systems depend on a valid field sampling scheme that specifies the location, timing and extent of sampling needed to provide robust detection for a given species and environmental context. This chapter focuses foremost on demonstrating, analyzing and refining such a field sampling scheme for non-native *Centrarchids* in Maine lakes.

Once shed, eDNA has a limited period of availability for detection that depends upon the environment. This window is influenced by physical, chemical and biological processes that degrade the DNA below fragment sizes that are identifiable, or that transport it into or out of the detection environment. In standing water, like lakes and ponds, DNA leaves the water column relatively quickly, 14-35 days, through degradation (e.g., UV light) and sedimentation (Dejean et al., 2011; Dunker et al., 2016). Water samples from surface and sub-surface waters of lakes and ponds thus generally contain eDNA of animals that were present within the past few weeks (Dejean et al., 2011; Eichmiller et al., 2014). Spatially, the concentration of eDNA likely depends primarily on where the target species are located, but also where the water moves it. Differences in wind, currents and seasonal mixing may affect where eDNA will be encountered (Takahara et al., 2012; Takahara et al., 2013; Eichmiller et al., 2014; Turner et al., 2015). Notably, prior studies of eDNA in aquatic systems have focused on relatively warm periods of the year when open water is easily accessible and fish are often most metabolically and behaviorally active. As such it is unknown how much eDNA is produced and how long it persists under colder winter conditions in regions with ice cover. However, lakes and ponds in Maine are typically under ice cover for more than a third of the year and sampling during winter ice season might be hypothesized to provide several benefits, including slower breakdown of eDNA, foot access to sampling sites in large bodies of water, and potentially less risk of false positive detections.

False positive eDNA detections arise where the DNA of a target species is detected, but the species is not truly present at the time of testing. One of the easier forms of false positives to control are those arising from accidental contamination of sampling or laboratory equipment with the target species' DNA. Risks of such contamination can be greatly reduced through rigorous decontamination procedures and use of procedural control samples that are able to signal when and where field samples are potentially affected (Willerslev & Cooper, 2005). A more challenging form of false positive to control can arise from actual presence of the target species eDNA due to unaccounted transport mechanisms, such as movements between waterways by predatory birds or dumping of ballast, bilge or live-well water from boats that travel between waterbodies (Willerslev et al., 2003; Thomsen et al., 2012a). There are no procedural control samples that can be performed to signal such false positives, leaving only decisions of when and where eDNA is sampled as a means to limit these false detections. Sampling of water during winter ice conditions could be effective in this respect given the absence of boats and fish-eating birds during this period of the year. However, this potential benefit must be weighed against associated factors that could increase risks of false negatives.

All monitoring technologies must be concerned with false negatives, or the probability that a species is present but not detected with the sampling process. Although the DNA amplification tools available for eDNA detection are very sensitive, they are not infallible. One form of false negative occurs from procedural errors in analyzing a water sample that actually did contain the target species eDNA, upon initial sampling. This could arise from physical and chemical processes that degrade the eDNA or inhibit its amplification (Willerslev & Cooper, 2005; Herder et al., 2014), or from processing errors during filtering, extraction or amplification of the sample. These false negatives can again be controlled by following rigorous sampling and processing procedures, using amplification chemistries that are robust to potential PCR inhibition, and by including positive controls that can signal such problems. Most positive controls for eDNA have been limited to inclusion of known positive samples of the target species, or surrogate DNA sequences, at the amplification stage. However, this does not fully account for potential failures at earlier stages of processing. One way to offset those earlier risks is to demonstrate

that it is possible to amplify the eDNA of other ‘non-target’ species from the same sample. This is the basis for development of the sunfish eDNA assay I described in the preceding chapter.

The other basis for false negatives relates to insufficient statistical power of a given field sampling design. eDNA of rare or common species can be at extremely low concentrations and patchy in space and time (Thomsen et al., 2012a; Deiner & Altermatt, 2014; Erickson et al., 2016). It is theoretically possible that eDNA of a species might be present at a given site at one point in time and not at another, be present at a given site and missed by a given water sample, or be present in a given sample but be missed during a given DNA extraction and amplification. These hierarchical error processes will determine the odds of detecting a given species when it is actually present in a waterbody, and thus determine the number of sites, samples and technical replicates required to attain a desired confidence in the absence of target species. Ideally, such a statistical assessment of sampling intensity is determined using data from a field assessment that corresponds to the intended future management uses of the tool. Hierarchical occupancy models use observed cases of detection or non-detection of target species to estimate parameters associated with both the presence of a species and the probabilistic sampling process used to detect it (Mackenzie et al., 2002; Royle et al., 2006). Although used most often with traditional observation surveys, these models have recently been applied to eDNA detection (Schmidt et al., 2013; Hunter et al., 2015). The resulting estimated probabilities of detection at the levels of sites within waterbodies, samples within sites, and qPCR replicates within samples can in turn be used to reduce risk of false negatives by establishing the amount and allocation of sampling effort needed to attain an overall probability of detecting a species when it is actually present.

Having developed lab assays and field sampling tools for detection of eDNA from three exotic *Centrarchids* (Chapter 2), my primary goal in this study was to provide a multi-season proof of utility for verifying previously confirmed or unconfirmed reports of illegal species introductions. My secondary goal was to use the data gathered from these field surveys to assess their statistical power and provide quantitative recommendations for future field survey designs. To meet these goals, I:

- 1) Conducted a field survey of 21 sites in Maine for eDNA detection of three *Centrarchid* species in winter (through the ice).
- 2) Conducted a field survey of the same 21 sites in Maine for eDNA detection of three *Centrarchid* species during the period in and around spring spawning (open water).
- 3) Compared detection patterns across seasons and MDIFW reports of different confidence of species presence.
- 4) Used hierarchical occupancy models to estimate the actual power of my surveys and project the power of alternative sampling schemes.

Methods

Study area

This study included eDNA sampling for 21 reported *Centrarchid* introductions in 16 Maine lakes (Figure 3.4). The lakes were chosen from a MDIFW database of confirmed and unconfirmed reports of exotic fish introductions. Based on this database, and local biologist knowledge, reports were selected to represent one of three confidence categories: strong populations (species is abundant and often a targeted fishery), newly confirmed populations (report confirmed by MDIFW staff), and unconfirmed angler reports. Each lake was sampled for at least one target species, though several were tested for multiple species in different categories, accounting for the difference between the number of lakes and number of reports for eDNA sampling (Table 3.1).

Table 3.1: Centrarchid eDNA survey sampling locations by lake, species and report category. Midas number refers to numeric system of lake designations used in Maine. Category corresponds to the category designated by MDIFW personnel or angler report status; ‘Strong’ being the target species is dominant within lake, ‘New’ being a newly confirmed target species (within 5 years) and ‘Unconfirmed’ being an angler report that has yet to be confirmed by IFW personnel. Species are Smallmouth Bass (SMB), Largemouth Bass (LMB) and Black Crappie (BC).

| Lake | Abbreviation | Midas | Area (km²) | Mean depth (m) | Deepest point | Category | Species |
|--------------|---------------------|--------------|------------------------------|-----------------------|----------------------|-----------------|----------------|
| Bog | BOG | 1258 | 3.4 | 5.2 | 14.0 | Strong | SMB |
| Fields | FIE | 4282 | 2.1 | 4.0 | 9.5 | Strong | BC |
| Eskutassis | ESK | 2250 | 3.6 | 3.4 | 12.2 | Strong | SMB |
| Spectacle | SPE | 5410 | 0.6 | 9.5 | 30.5 | Strong | LMB |
| Webber | WEB | 4857 | 5.0 | 5.5 | 12.5 | Strong | LMB |
| Brewer | BRE | 4284 | 3.9 | 7.9 | 14.6 | New | SMB |
| Bog | BOG | 1258 | 3.4 | 5.2 | 14.0 | New | LMB |
| Chemo | CHE | 4278 | 5.0 | 4.0 | 7.3 | New | BC |
| Eskutassis | ESK | 2250 | 3.6 | 3.4 | 12.2 | New | LMB |
| Kimball | KIM | 5330 | 0.2 | 3.0 | 5.8 | New | SMB |
| Mattanawcook | MAT | 2226 | 3.4 | 2.7 | 6.7 | New | BC |
| Sabattus | SAB | 3796 | 8.0 | 4.2 | 5.8 | New | BC |
| Webber | WEB | 4857 | 5.0 | 5.5 | 12.5 | New | BC |
| China | CHI | 5448 | 16.0 | 8.5 | 26.0 | Unconfirmed | BC |
| Davis | DAV | 4276 | 2.0 | 3.0 | 4.2 | Unconfirmed | BC |
| Graham | GRA | 4350 | 38.0 | 5.2 | 14.3 | Unconfirmed | BC |
| Graham | GRA | 4350 | 38.0 | 5.2 | 14.3 | Unconfirmed | LMB |
| Pleasant | PLE | 1590 | 1.4 | 5.2 | 11.0 | Unconfirmed | LMB |
| Scraggly | SCR | 9649 | 6.6 | 6.7 | 12.8 | Unconfirmed | LMB |
| Spectacle | SPE | 5410 | 0.6 | 9.5 | 30.5 | Unconfirmed | BC |
| Sysladobsis | SYS | 4730 | 22.0 | 7.6 | 20.1 | Unconfirmed | LMB |

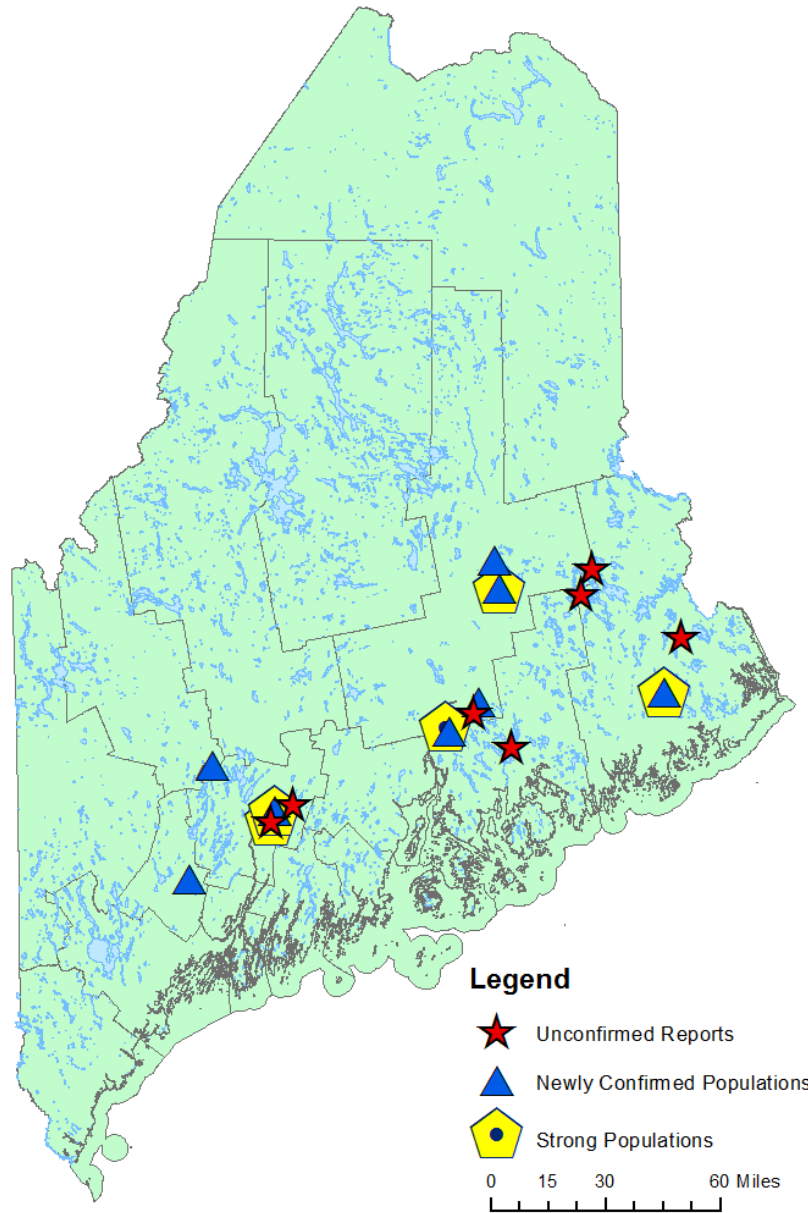


Figure 3.1: Centrarchid eDNA survey sampling locations by category. Table 3.1 lists location names.

eDNA water sampling and pre-processing

Each lake was sampled from three sites. The sites were distributed across multiple regions of each lake where possible, targeting species-preferred habitats during the season sampled. Bass and Black Crappie inhabit deep water regions of lakes (e.g., 6-10 m) in the winter but spawn and feed in shallower

areas during spring/summer, and thus winter and spring/summer sampling sites rarely overlapped. Using depth maps I selected the locations I believed the target species would be present during winter and spring. During field surveys, I confirmed depths at those locations or adjusted my sampling location to account for map inaccuracies. At each site three 1 L water samples were taken 10-20m apart. Hence, 9 L of water were sampled from each lake during each sampling session. The specific manner of water collection was season dependent, as outlined below.

Each 1 L water sample was collected into two 500 ml commercial drinking water bottles (Nestle PureLife), which were disposed of and never reused following extraction in order to avoid the potential for contamination. Samples were kept on ice until filtering. All samples were filtered the same day through 1.5 micron Whatman glass fiber filters (5.5 cm diameter) using a vacuum pump. The filters were stored at -20⁰ C until extraction, which occurred within 5 days of sampling. DNA was extracted from the filters using Qiagen blood/tissue kits (Qiagen, Valencia, CA). Extracted DNA was stored at -20⁰ C until testing with the species-specific eDNA assays, which occurred within six weeks of the initial collection.

Winter sampling

Winter sampling employed the deep-water sampling system described in Chapter 2. Briefly, ice was cut with a standard battery powered hand drill and 2.54 cm diameter wood auger drill bit. The deep-water sampler, constructed from PVC tubing, was rigged to sample water from 1-1.5 meters off the lake bottom and deployed down these holes. Water retrieved with the sampler was then poured into the collection bottles. The volume of the water sampler was approximately 650 ml, so the sampler was repeatedly deployed to obtain a full sample. The drill bit and sampler were in turn sterilized between lakes by fully submerging them in 10% bleach solution for a minimum of 10 minutes, then rinsing thoroughly with sterile bottled water.

Sampling locations were targeted as locations with depths between 6 and 10 m, where possible, using MDIFW lake depth maps, and depths were subsequently confirmed on site. However, late season ice conditions constrained site selection in some cases, and some lakes had maximum depths of less than

6 m. All sampling was conducted between 11 February and 4 March 2016, a period when all lakes were ice locked and safe to travel on and before full ice out in late winter.

Spring sampling

All three target species begin spawning in the spring when water temperatures reach at least 15.5⁰ C (Jordan, 2001; Lucas, 2002). Spring sampling locations were selected to target near shore habitats that fish occupy during the pre-spawn, spawning and post-spawning window. These habitats are typified by relatively shallow water (0.5-3.0 meters), early season macrophyte growth, overhanging trees and fallen limbs. Sampling sites were again distributed in multiple regions of a given lake to increase odds of detection of species that might not be fully established in all preferred habitats. Sampling was conducted when the near shore water had reached 18.5-21⁰ C and was conducted between 24 May and 11 June 2016. Surface water samples were collected directly into new commercial drinking water bottles (Nestle PureLife) by dipping each empty bottle into the lake. All field equipment, including footwear and transport coolers were sterilized between lakes by washing with 10% bleach solution and then rinsing with DI water.

In order to test any effect the winter sampling device may have made on DNA collection, I conducted duplicate spring sampling with the winter and spring sampling procedures at four lakes (Chemo Pond, Davis Pond, Fields Pond and Webber Pond). Unlike my actual winter sampling that occurred at depth, however, the winter water sampler was deployed at the surface at the exact same location and date where I sampled surface water directly into bottles for my standard spring sampling. The resulting paired samples were then compared for eDNA concentrations (Appendix C).

Summer sampling

Two lakes with low and inconsistent levels of target species detection in winter or spring were resampled in the summer to provide further potential confirmation. The two lakes re-sampled were China Lake and Sysladobsis Lake. These lakes were sampled on 1 and 2 July 2016. The lakes were sampled

from three near-shore locations. I selected new sites for their distance from boat launches (minimum 3.5 km) as well as from the previous collection sites in an attempt to limit potential contamination from off-site sources, as well as achieve a wider survey of the lake. Water was collected as in the spring, directly into sample bottles from water surface.

qPCR

Each lake's extracted samples were run on a single 96-well qPCR plate on a Bio-Rad CFX96 Real-Time System thermal cycler using the optimized primer and thermal conditions described in the preceding chapter. Each extracted water sample (1 L volume) was run with four technical replicates per plate. Reactions were 20 μ l in total volume, 10 μ l of Taqman Environmental Master Mix 2.0 (Applied Biosystems), 2.6 μ l of primer-probe, 4.4 μ l of nuclease free water and 3 μ l of template. Primer-probe concentrations were 10 μ M/5 μ M for all assays. The thermal cycle protocol was an initial denaturation at 95⁰ C for 10 minutes followed by 50 cycles of denaturation at 95⁰ C for 10 seconds, annealing and extension at 60⁰ C for 30 seconds. Any well with C_q value prior to cycle 45 was considered a positive amplification of target DNA (Wilcox et al., 2013). The baseline threshold was automatically calculated by the thermal cycler software between cycles 2 and 20. Any lake-species combination that had no amplification on the first qPCR was run on a second identical plate. Following qPCR, all plates were archived at 4⁰ C in case of a need for follow-up analyses, including sequencing (see below).

Controls

At each lake, three 1 L cooler blank samples, 1 per sampling location, were collected during field sampling by opening the bottle into the air and then recapping. The cooler blanks were transported and processed alongside the field water. A filtering control was created for each site by filtering 500 ml of deionized lab water through the same filter and apparatus used for field samples prior to filtering samples from a given site. No-template controls were run on each plate by preparing wells with all necessary

qPCR chemistry but substituting DNA free water in place of an extracted sample. The above listed controls were all run with four technical replicates.

In addition to the assays for the target species, one water sample from each site was run with four technical replicates using a sunfish primer-probe set that joint amplifies Redbreast Sunfish (*Lepomis auritus*) and Pumpkinseed Sunfish (*Lepomis gibbosus*) DNA as a positive field water control (see Chapter 2). In addition, Taqman Internal Exogenous Control (Applied Biosystems) was used in these same wells as an additional internal positive control (IPC) in order to test for PCR inhibition. This multiplexing was made possible because the Sunfish and IPC probes were labelled with different fluorophores, 6-FAM and VIC, respectively, and because the IPC control is designed by the manufacturer to not compete with target DNA during amplification. Each 20 µl reaction had 2 µl of 10x IPC mix and 0.4 µl of 50x IPC DNA. Synthetic gene fragments were run with six five-fold dilutions (31250, 6250, 1250, 250, 50, 10 copies/µl) on every plate to ensure all reagents and protocol were functioning optimally and were used to estimate qPCR efficiency via standard curve analysis.

Sequencing

In order to confirm positive detections of the target species in lakes where the species was previously unconfirmed by MDIFW, I selected two positive samples (PCR plate wells) for DNA sequencing. I purified the PCR products using Purelink PCR purification kit (ThermoFisher Scientific) and the samples were sequenced at the University of Maine DNA sequencing facility. I used BLAST (Basic Local Alignment Search Tool; Genbank, www.ncbi.nlm.nih.gov/blast) to confirm the identity of the sequences against online databases.

Analysis

I viewed qPCR results using Bio-Rad CFX Manager 3.1 (Bio-Rad Laboratories). I checked each amplification curve for appropriate shape and reasonable cycle start point (23-45). I exported the C_q

values, the estimated cycle number where fluorescence climbs above the baseline threshold, from the software, disregarding C_q values that were likely to represent analytical error on the basis of being unrealistic (i.e. those not between 23-45 cycles) or showing inappropriately shaped curves (e.g. spiked, multiple curves in a single well).

Occupancy Models

I estimated the probability of detection for each lake with positive eDNA amplification using 3-tiered hierarchical occupancy models. The models I used are top-down models, where each level's occupancy is dependent on the next higher level's and are described in Mordecai (2011) and Kery & Royle (2016) as:

$$\begin{array}{ll}
 \text{Site presence/absence:} & z_i: z_i \sim \text{Bernoulli}(\psi) \\
 \text{Sample presence/absence:} & a_{ij}: a_{ij}|z_i \sim \text{Bernoulli}(z_i * \theta) \\
 \text{Replicate presence/absence:} & y_{ijk}: a_{ijk}|a_{ij} \sim \text{Bernoulli}(a_{ij} * p)
 \end{array}$$

z_i and a_{ij} are random variables of presence/absence for site i and sample j . y_{ijk} is the detection/non-detection for replicate k , sample j and site i . ψ is the unit level occupancy probability (site), which is the probability that target eDNA was present at a given site in the lake. θ is the next lower level occupancy probability (sample), the probability that eDNA was collected in a sample taken at site i . p is the detection level occupancy probability (qPCR replicate), the probability, that DNA was amplified in qPCR replicate k , given that eDNA was present at the site and collected in a sample. Utilizing p , I calculated the cumulative probability (p^*) that I amplified target DNA in K replicates, $p^* = 1 - (1-p)^K$. In this way I determined the power of my lab analysis design to detect eDNA in a given sample and to project how changing the number of technical replicates would influence this power. Similarly, I calculated the cumulative probability that eDNA was collected in J samples (θ^*), $\theta^* = 1 - (1-\theta)^J$, in order to determine the power of my actual sample replication, and to project how changes in sample replication would influence the probability of detecting target eDNA at a given site. Finally, I calculated the

cumulative probability of collecting eDNA in a given lake when sampling I sites, $= 1-(1-\psi)^I$. The models were run through WinBUGS 1.4.3 software (Lunn et al., 2000) via R software program version 3.3.1 (www.r-project.org). Code adapted from Kery & Royle (2016) available in Appendix D.

Results

Table 3.2: Exotic Centrarchid eDNA survey detections summarized by number of positive sampling sites. Lake and species are categorized by MDIFW report confidence: Strong = well established population/fishery, New = species presence recently confirmed by MDIFW staff, Unconfirmed = angler report that was not yet confirmed by MDIFW staff at the time of sampling. Positive (Y) or Negative (N) detection outcomes are tabulated along with the number of lake sites (out of a maximum three per season) showing at least one positive amplification of the target species.

| Lake | Midas | Category | Species | Winter (Positives) | Spring (Positives) | Summer (Positives) |
|--------------|-------|-------------|---------|-----------------------|-----------------------|-----------------------|
| Bog | 1258 | Strong | SMB | Y (1) | Y (3) | |
| Fields | 4282 | Strong | BC | Y (3) | Y (3) | |
| Eskutassis | 2250 | Strong | SMB | Y (2) | Y (3) | |
| Spectacle | 5410 | Strong | LMB | Y (3) | Y (3) | |
| Webber | 4857 | Strong | LMB | Y (2) | Y (3) | |
| Brewer | 4284 | New | SMB | Y (3) | Y (3) | |
| Bog | 1258 | New | LMB | N (0) | Y (2) | |
| Chemo | 4278 | New | BC | Y (1) | Y (3) | |
| Eskutassis | 2250 | New | LMB | Y (1) | Y (3) | |
| Kimball | 5330 | New | SMB | N (0) | N (0) | |
| Mattanawcook | 2226 | New | BC | Y (3) | Y (2) | |
| Sabattus | 3796 | New | BC | Y (2) | Y (3) | |
| Webber | 4857 | New | BC | Y (2) | Y (3) | |
| China | 5448 | Unconfirmed | BC | N (0) | Y (1) | N (0) |
| Davis | 4276 | Unconfirmed | BC | Y (1) | Y (3) | |
| Graham | 4350 | Unconfirmed | BC | Y (1) | Y (2) | |
| Graham | 4350 | Unconfirmed | LMB | Y (2) | Y (3) | |
| Pleasant | 1590 | Unconfirmed | LMB | Y (1) | Y (2) | |
| Scraggly | 9649 | Unconfirmed | LMB | N (0) | N (0) | |
| Spectacle | 5410 | Unconfirmed | BC | N (0) | N (0) | |
| Sysladobsis | 4730 | Unconfirmed | LMB | Y (1) | Y (2) | Y (2) |

Winter sampling

Detections from all waterbodies and seasons of sampling are enumerated with respect to number of positive sampling sites in Table 3.2. Samples from all introductions within the strong/well established category produced positive amplifications of target species DNA in at least a subset of sites, samples and qPCR replicates. Six of eight newly confirmed introductions had positive amplifications. Four of eight unconfirmed reports had clear positive amplifications. One unconfirmed site, Sysladobsis Lake, had a questionable amplification at the sampling site near the boat launch in winter. There was an amplification curve, but it plateaued very early, and appeared similar to the IPC curves, resulting in an extremely high estimated eDNA copy number. Based on these observations, I considered this amplification suspect. Of the 21 lakes, 10 had no amplification on the initial plate and were run on a second identical plate. Four of those resulted in positive amplifications on the second plate, Chemo Pond, Davis Pond, Graham Lake (BC) and Pleasant Lake.

Spring sampling

All putative introductions within the strong/well established category showed positive amplification of target species DNA. All but one newly confirmed introduction showed positive amplification for target species, one more than had in winter. A single reported introduction in the newly confirmed introduction was negative in spring and winter, Kimball Pond, Vienna, Maine. Five of eight unconfirmed reports had clear positive amplifications, including Sysladobsis Lake. One unconfirmed site, China Lake, was negative in winter, but had a single positive amplification in spring, representing a single positive qPCR reaction out of 72 tested, however that positive amplification was from a sample collected near (>100 m) the public boat launch.

Summer sampling

I resurveyed the two lakes with different or questionable detections in winter and spring, during summer. Sysladobsis Lake was resampled at three new independent locations. Two of the three locations had positive amplification of target species DNA. China Lake was similarly resampled, but none of the summer samples provided positive amplification of target species DNA.

Standard curves and starting copy number

Using the synthetic gene calibration curves relating gene copy numbers and fluorescence, I estimated the eDNA concentrations in my reactions and in turn the eDNA copy number in my positive, 1 L field samples. The average estimated starting copy number for winter samples was 355 copies / L sampled water. The winter top and bottom ten estimated copy numbers in qPCR reactions for positive samples are summarized in Table 3.3, and covered a range from approximately 12.3 to 2901.8 copies per sample.

Table 3.3: Winter top and bottom ten estimated starting copy number / 1 L sample. Estimates made via standard curve analysis.

| Lake | Content | Category | Estimated copy number |
|--------------|---------|-----------------|-----------------------|
| Bog | SMB | Strong | 2902 |
| Sabattus | BC | Newly Confirmed | 2444 |
| Bog | SMB | Strong | 719 |
| Chemo | BC | Newly Confirmed | 515 |
| Webber | LMB | Strong | 423 |
| Brewer | SMB | Newly Confirmed | 394 |
| Eskutassis | SMB | Strong | 348 |
| Eskutassis | LMB | Newly Confirmed | 304 |
| Webber | BC | Newly Confirmed | 281 |
| Mattanawcook | BC | Newly Confirmed | 279 |

Table 3.3: Continued

| | | | |
|--------------|-----|-----------------|----|
| Graham | LMB | Newly Confirmed | 44 |
| Brewer | SMB | Unconfirmed | 41 |
| Graham | LMB | Newly Confirmed | 38 |
| Mattanawcook | BC | Strong | 37 |
| Fields | BC | Strong | 36 |
| Spectacle | LMB | Unconfirmed | 34 |
| Pleasant | LMB | Unconfirmed | 34 |
| Graham | LMB | Strong | 33 |
| Fields | BC | Strong | 16 |
| Fields | BC | Unconfirmed | 12 |

The average estimated starting copy number for spring samples was 4598 copies/ L of sampled water. The spring top and bottom ten estimated starting copy numbers are summarized in Table 2.4 and ranged from 12 to 96,620 copies/L. Comparing the mean estimated copy numbers in positive samples by lake, there was support for significantly higher starting copy number in the spring compared to winter (paired two sample t-test: $T = 15.327$, $d.f. = 935$, $p\text{-value} = <0.001$).

Table 3.4: Spring top and bottom ten estimated starting amplifiable eDNA copy number / 1 L sample. Estimates made via standard curve analysis.

| Lake | Content | Category | Estimated copy number |
|--------|---------|----------|-----------------------|
| Webber | BC | New | 96620 |
| Webber | BC | New | 44294 |
| Webber | LMB | Strong | 38132 |
| Webber | LMB | Strong | 21701 |
| Webber | LMB | Strong | 20685 |
| Webber | BC | New | 19500 |
| Webber | LMB | Strong | 17886 |
| Webber | LMB | Strong | 16525 |

Table 3.4: Continued

| | | | |
|-------------|-----|-----------------|-------|
| Webber | BC | New | 14489 |
| Webber | BC | New | 11954 |
| Pleasant | LMB | Unconfirmed | 47 |
| Sysladobsis | LMB | Unconfirmed | 47 |
| Eskutassis | SMB | Strong | 39 |
| Eskutassis | LMB | Newly Confirmed | 35 |
| Chemo | BC | Newly Confirmed | 27 |
| Chemo | BC | Newly Confirmed | 22 |
| Graham | BC | Unconfirmed | 20 |
| Bog | LMB | Newly Confirmed | 13 |
| Graham | BC | Unconfirmed | 13 |
| Graham | BC | Unconfirmed | 12 |

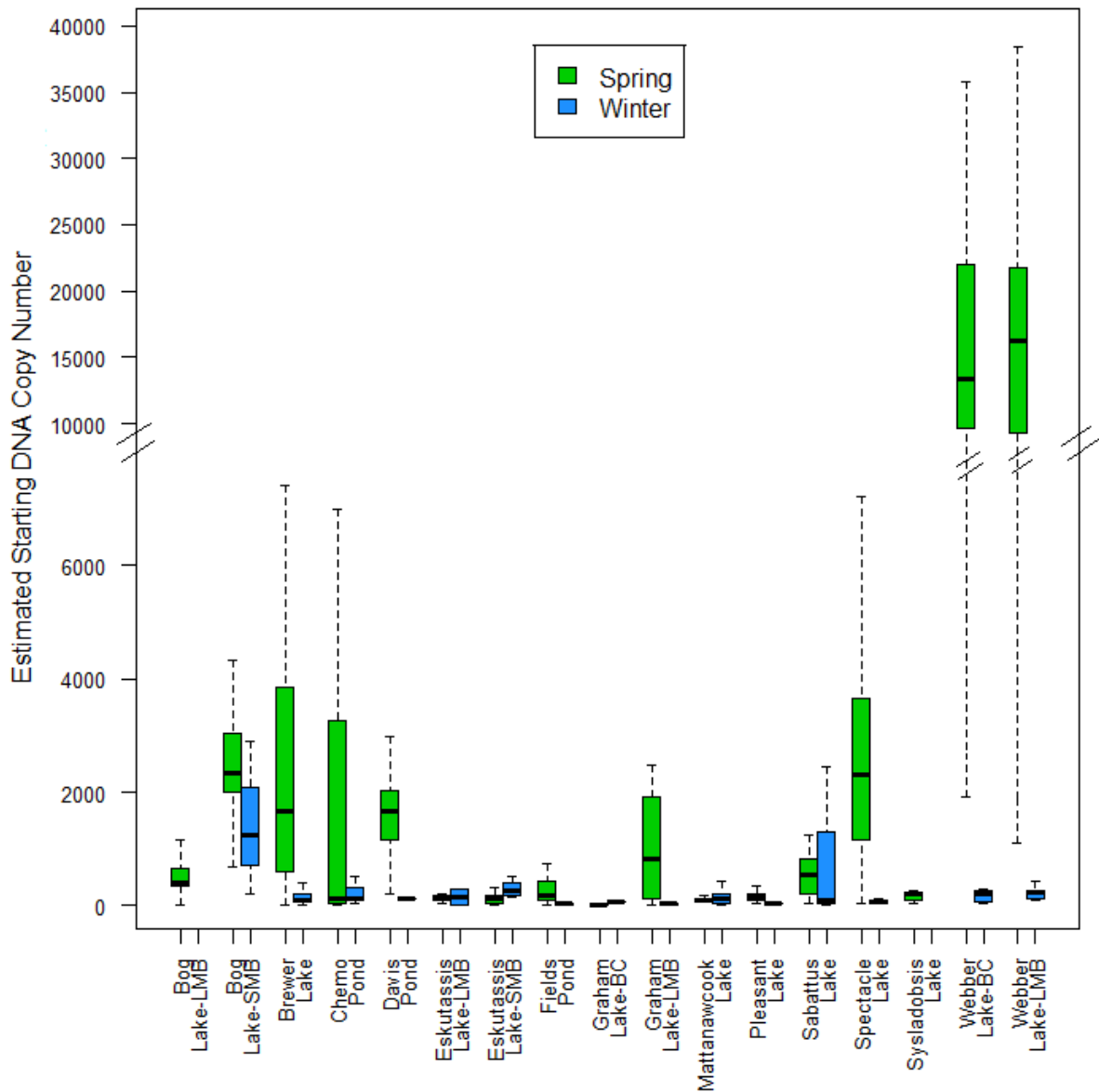


Figure 3.2: Estimated starting amplifiable eDNA copy number / 1L water sample by lake and season. Boxes represent 25th and 75th quartiles, black band represents 50th quartile. Whiskers represent full range of values excluding outliers. Green boxes are spring estimates, blue boxes are winter estimates.

Sequencing

Two positive wells from each unconfirmed introduction were sequenced at the UMaine DNA sequencing facility. BLAST results matched 99-100% for target species in all but one plate. The initial

Sysladobsis Lake positive from winter was a poor match (66%) for the Largemouth Bass target species, however the top returns were all *Centrarchid* species with Largemouth being the best fit, suggesting this could have been due to a poor sequencing run associated with low DNA concentration.

Table 3.5: Sequencing results from unconfirmed sites. Sequences of PCR products from two wells per previously unconfirmed site. All sequences high percent match for target species except Sysladobsis spring (WA12).

| Lake/Well | Species Targeted | Species Match | Percent Match |
|------------------|-------------------------|----------------------|----------------------|
| Davis/A1 | Black Crappie | Black Crappie | 99 |
| Davis/B4 | Black Crappie | Black Crappie | 99 |
| Graham/B10 | Black Crappie | Black Crappie | 99 |
| Graham/C1 | Black Crappie | Black Crappie | 99 |
| Graham/B6 | Largemouth Bass | Largemouth Bass | 99 |
| Graham/A2 | Largemouth Bass | Largemouth Bass | 100 |
| Pleasant/A5 | Largemouth Bass | Largemouth Bass | 99 |
| Pleasant/C6 | Largemouth Bass | Largemouth Bass | 100 |
| Sysladobsis/A9 | Largemouth Bass | Largemouth Bass | 99 |
| Sysladobsis/B8 | Largemouth Bass | Largemouth Bass | 99 |
| Sysladobsis/WA12 | Largemouth Bass | Largemouth Bass | 66 |

Controls

No field, equipment or non-template controls showed any amplification. Synthetic gene fragments amplified generally as expected along standard curves, though there were 6 wells in total (out of 342) with no amplification. Sunfish positive controls amplified at least one sample in all sites except Sabattus Lake during winter. Means of the winter and spring IPC C_q values were 27.1 and 27.2, and standard deviations were 1.78 and 1.92, respectively. A paired t-test comparing IPC C_q values across seasons was non-significant ($T = -0.7056$, $d.f. = 596.35$, $p = 0.48$), indicating that there was not a difference between the two seasons. Tukey's tests across surveys within seasons indicated that two surveys had higher C_q values than the remainder in winter and the same two had higher values in spring (Fields Pond and Graham Lake). These larger C_q values suggest potential PCR inhibition in these waters, but this did not appear to influence detection in those cases as they were both positive for their reported target species.

Occupancy Models

Table 3.6: Winter occupancy model probability estimates. Posterior mean and 95 % confidence intervals (95% CI) for all lakes with positive amplification of target eDNA. ψ is the average probability of eDNA presence at site within lake. θ is the average probability of collecting eDNA in a single sample. p is the average probability estimate of detecting eDNA in a qPCR replicate.

| Lake | Species | ψ | 95% CI | θ | 95% CI | p | 95% CI |
|------|---------|--------|-----------|----------|-----------|-------|-----------|
| BOG | SMB | 0.448 | .077-.907 | 0.534 | .013-.985 | 0.426 | .008-.990 |
| BRE | SMB | 0.801 | .398-.994 | 0.801 | .390-.994 | 0.61 | .192-.936 |
| CHE | BC | 0.472 | .081-.921 | 0.409 | .011-.983 | 0.413 | .012-.988 |
| DAV | BC | 0.235 | .006-.750 | 0.472 | .002-.973 | 0.498 | .024-.977 |
| ESK | LMB | 0.684 | .224-.985 | 0.456 | .009-.967 | 0.313 | .007-.868 |
| ESK | SMB | 0.657 | .206-.981 | 0.498 | .009-.989 | 0.422 | .059-.924 |
| FIE | BC | 0.801 | .397-.994 | 0.491 | .007-.994 | 0.395 | .047-.936 |
| GRA | BC | 0.474 | .076-.936 | 0.41 | .010-.982 | 0.37 | .011-.987 |
| GRA | LMB | 0.647 | .205-.977 | 0.598 | .102-.990 | 0.348 | .005-.933 |
| MAT | BC | 0.797 | .407-.993 | 0.8 | .407-.993 | 0.385 | .120-.817 |
| PLE | LMB | 0.481 | .078-.940 | 0.41 | .012-.983 | 0.369 | .011-.986 |
| SAB | BC | 0.668 | .223-.982 | 0.53 | .095-.960 | 0.35 | .006-.905 |
| SPE | LMB | 0.802 | .402-.994 | 0.631 | .071-.994 | 0.387 | .037-.956 |
| WEB | BC | 0.664 | .221-.982 | 0.502 | .089-.942 | 0.47 | .062-.918 |
| WEB | LMB | 0.634 | .207-.969 | 0.669 | .105-.992 | 0.343 | .004-.952 |

Table 3.7: Spring occupancy model probability estimates. Posterior mean and 95 % confidence intervals (95% CI) for all lakes with positive amplification of target eDNA. ψ is the average probability of eDNA presence at site within lake. θ is the average probability of collecting eDNA in a single sample. p is the average probability estimate of detecting eDNA in a qPCR replicate.

| Lake | Species | ψ | 95% CI | θ | 95% CI | p | 95% CI |
|------|---------|--------|-----------|----------|-----------|-------|-----------|
| BOG | LMB | 0.631 | .204-.967 | 0.569 | .086-.991 | 0.704 | .281-.995 |
| BOG | SMB | 0.803 | .393-.995 | 0.802 | .397-.994 | 0.908 | .685-.998 |
| BRE | SMB | 0.801 | .413-.994 | 0.803 | .409-.994 | 0.864 | .559-.998 |
| CHE | BC | 0.804 | .409-.994 | 0.802 | .393-.994 | 0.795 | .343-.998 |
| DAV | BC | 0.611 | .190-.951 | 0.656 | .095-.991 | 0.858 | .540-.996 |
| ESK | LMB | 0.802 | .398-.994 | 0.563 | .071-.969 | 0.47 | .195-.857 |
| ESK | SMB | 0.798 | .402-.994 | 0.8 | .388-.994 | 0.749 | .440-.975 |
| FIE | BC | 0.798 | .401-.994 | 0.802 | .400-.995 | 0.865 | .438-.998 |
| GRA | BC | 0.659 | .227-.980 | 0.519 | .095-.958 | 0.407 | .053-.913 |
| GRA | LMB | 0.8 | .390-.994 | 0.8 | .162-.994 | 0.819 | .442-.997 |
| MAT | BC | 0.652 | .211-.979 | 0.493 | .008-.989 | 0.481 | .064-.929 |

Table 3.7: Continued

| | | | | | | | |
|-----|-----|-------|-----------|-------|-----------|-------|-----------|
| PLE | LMB | 0.802 | .402-.993 | 0.428 | .071-.967 | 0.469 | .058-.920 |
| SAB | BC | 0.795 | .380-.994 | 0.802 | .396-.994 | 0.864 | .452-.998 |
| SPE | LMB | 0.803 | .393-.995 | 0.802 | .395-.994 | 0.908 | .685-.998 |
| SYS | LMB | 0.655 | .214-.978 | 0.491 | .087-.927 | 0.312 | .006-.798 |
| SYS | LMB | 0.469 | .081-.925 | 0.542 | .011-.986 | 0.353 | .088-.899 |
| WEB | BC | 0.8 | .401-.994 | 0.734 | .204-.993 | 0.851 | .089-.998 |
| WEB | LMB | 0.803 | .393-.995 | 0.802 | .395-.994 | 0.908 | .084-.998 |

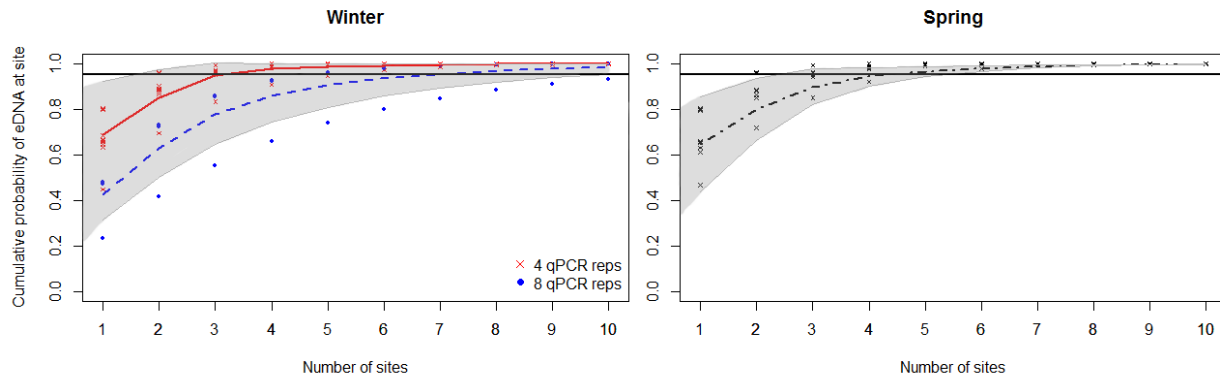


Figure 3.3: Estimated cumulative probability of target eDNA presence at sites in a lake. Each point represents a survey (lake and species) specific estimate for a positive detection within the original confirmed or unconfirmed categories. The left panel is for surveys performed in winter and the right panel is for surveys conducted in spring. The fit lines are the mean probabilities across surveys and the gray shaded area is the 95% confidence bound for all surveys. In winter two lines are fit, one for the samples where only 4 qPCR replicates were required for detection and one where 8 qPCR replicates were performed (rerun of plate). The black horizontal line represents a 95% probability of eDNA at the site.

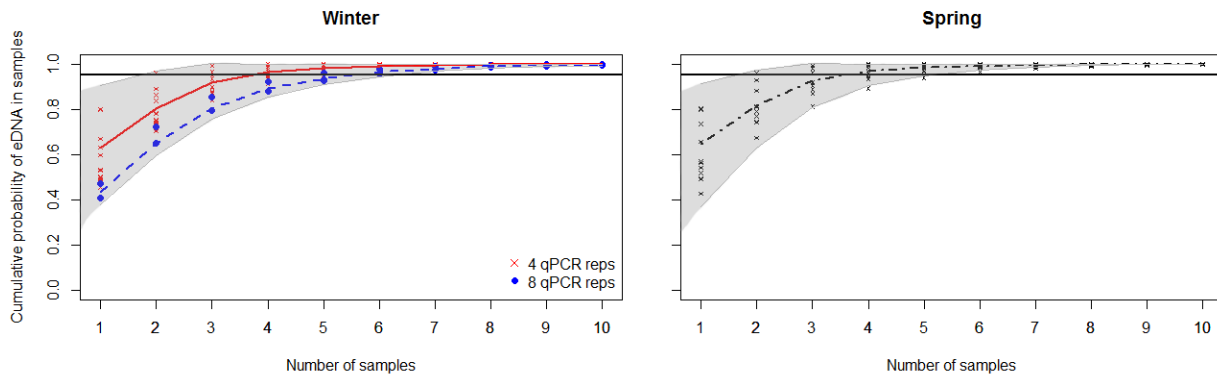


Figure 3.4: Estimated cumulative probability of collecting target eDNA in samples. Each point represents a survey (lake and species) specific estimate for a positive detection within the original confirmed or unconfirmed categories. The left panel is for surveys performed in winter and the right panel is for surveys conducted in spring. The fit lines are the mean probabilities across surveys and the gray shaded area is the 95% confidence bound for all surveys. In winter two lines are fit, one for the samples where only 4 qPCR replicates were required for detection and one where 8 qPCR replicates were performed (rerun of plate). The black horizontal line represents a 95% probability of eDNA at the site.

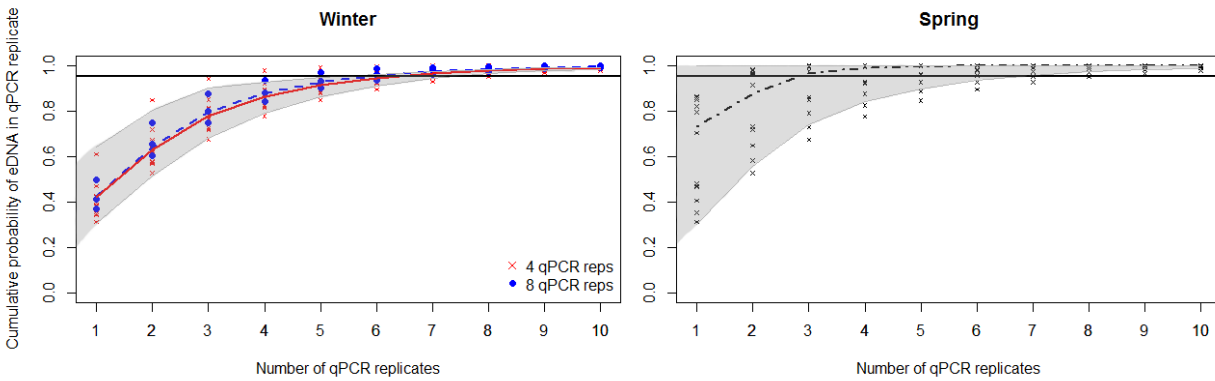


Figure 3.5: Estimated cumulative probability of detecting target eDNA in a qPCR replicates. Each point represents a survey (lake and species) specific estimate for a positive detection within the original confirmed or unconfirmed categories. The left panel is for surveys performed in winter and the right panel is for surveys conducted in spring. The fit lines are the mean probabilities across surveys and the gray shaded area is the 95% confidence bound for all surveys. In winter two lines are fit, one for the samples where only 4 qPCR replicates were required for detection and one where 8 qPCR replicates were performed (rerun of plate). The black horizontal line represents a 95% probability of eDNA at the site.

Plotting the cumulative probability curves for all positive surveys at the site, sample and qPCR level showed that I had reasonably high cumulative probability (>75% to >97%) of collecting eDNA when sampling three sites per lake, collecting eDNA in a sample when taking three samples at a site, and of amplifying eDNA with four qPCR replicates per sample. However, there were some surveys with lower estimated detection probabilities, including those surveys in which the target species was only detected with a second qPCR plate that brought the total replicates to eight instead of four. However, even in these cases cumulative probabilities were estimated in the range of 54% to >67% suggesting detection was still in my favor. Ultimately, the lower ranges on detection probabilities were similar across seasons for a given number of samples or qPCR replicates, consistent with functional limits of detection for low concentration eDNA samples regardless of season. However, mean detection probabilities were generally higher in spring than winter at all levels (by 10-30%), and the lower range of detections was markedly higher (by 25%) at the site level in spring than winter indicating fewer sites may need to be surveyed in spring than in winter for the same level of confidence.

Discussion

In this study I surveyed 16 Maine lakes during two seasons for 21 reported introductions of non-native *Centrarchids* using newly developed eDNA assays. The qPCR assays were highly effective at detecting my target *Centrarchid* species, Largemouth Bass, Smallmouth Bass and Black Crappie. I further put to full use my winter sampling protocol and device for through-ice collections and proved its efficacy at collecting eDNA. I identified the presence of exotic species in all lakes where they were reported to be abundant, in most lakes where they were newly confirmed, and in five lakes where they were only suspected from unverified angler reports. Although there was a significant difference in eDNA concentrations, and number of positive samples collected between winter and spring seasons, this translated into only modest differences in overall presence detection with the particular sampling design I employed. I ran occupancy models for three levels of sampling detection using my data and determined that this modest difference between seasons was due to my having fairly high odds of species detection

overall, except in sites with the lowest eDNA and presumptive species abundance. In the remaining of the discussion I consider the relative utility and efficacy of winter and spring sampling, how my sampling design compares to others used for eDNA detection, and make recommendations for future eDNA sampling designs based on my quantitative estimates of detection probabilities and the need to control overall rates of false negatives.

When considering that a single positive amplification constituted detection of target species DNA, there was relatively little difference between the two seasons in overall inferred presence or absence at each location. Given that spring and summer sampling are the norm, this is perhaps made clearest by examining how winter sampling results differed. There was a much higher estimated starting DNA copy number in spring than in winter, suggesting that the fish are more active or that the DNA is more available to sample; however, the higher concentration of DNA didn't make an overall difference for detection at the lake level. All surveys that were negative in spring were also negative in winter. Likewise, the vast majority of surveys that were positive in spring were also positive in winter, suggesting real presence of the species. Nominally, exceptions to this included Largemouth Bass in Bog lake, Black Crappie in China Lake, and Largemouth Bass in Sysladobsis. However, a subsequent negative summer survey makes it uncertain that the single spring positive detection in China Lake was real, and the assumption of a negative detection in winter for Sysladobsis only holds if the one very questionable winter positive there is discounted. Even the case of negative Largemouth Bass detection in Bog Lake, a newly confirmed introduction, may have more to do with field sampling constraints than reduced winter detection capabilities.

When sampling for eDNA, targeting where the fish are located within water bodies increases the likelihood of positive detection (Eichmiller et al., 2014), particularly if there is little water movement to redistribute eDNA elsewhere. This may be particularly important in the winter, when my target species are much less active and widespread under the ice. The ice conditions on Bog Lake did not allow for foot access across the whole lake and several roads that access the lake were unplowed when I was sampling, restricting my sampling to only a single basin. During the spring, I was able to sample the lake more

broadly and amplified both LMB and SMB as well as SUN, as expected. This result emphasizes that foot access over ice to sampling sites can be both an advantage and constraint on winter sampling.

However, spring and summer surveys are not without their own logistical constraints. China Lake, is a large lake near a relatively developed area in Southern Maine. It is a very popular lake with boaters and anglers. I found no amplification of the target species, Black Crappie, in winter. However, in the spring I amplified target eDNA in one sample, from a location near one of the public boat launches. I resampled the lake at three new locations, careful to choose spots as far from public access points as possible, and had no further amplification. Hence, the single amplification could have been a false positive associated boat traffic carrying eDNA from another location with BC. Surveying in winter likely greatly reduces the risk of this sort of potential contamination. However, I cannot conclude that this single detection was a false positive due to boats. Given there was a prior, unconfirmed, angler report for Black Crappie in this lake, and that the Lake is quite large, at 1.6 km² in area with a maximum depth of 26 m, there is certainly potential for a small population of Black Crappie and further eDNA surveys are recommended here. Nonetheless, this does not undermine the key point that the open water season is associated with greater uncertainty of potential false positives from eDNA transport due to anglers, birds, runoff, or other sources, than winter sampling.

Although winter and spring sampling provide fairly consistent species detections at the overall lake level, there were clear differences between the two with respect to numbers of sites where the species were detected. Winter surveys frequently resulted in one or two (out of three total) fewer sites with detections than spring surveys. Again, this might be anticipated on the grounds that my target species tend to concentrate in specific deep water locations in the winter but are more widely ranging in spring and summer (Peat et al., 2016). Likewise, lake waters are much more extensively mixed by wind and other processes in the spring than during the winter ice period. This difference in number of sites where eDNA is likely to be detected could be considered a plus or minus for confirming angler reports of fish introductions. On the one hand this patchy distribution of the fish and their eDNA can reduce the odds of encountering the target species if too few sites are sampled in a survey without any additional

information. On the other hand, if an angler reports a capture of a species from a particular part of a lake in winter, there are good odds that subsequent water sampling at that site will turn up that species' eDNA. In contrast, there are greater odds in spring and summer that a few individuals of a target species might have since moved elsewhere or their eDNA diluted by water mixing reducing detection success. If the goal, however, is to apply eDNA for pro-active detection in waters without reports, rather than to confirm angler reports, there may be a general detection advantage to spring surveys, given the relative ease of surface water sampling at more sites, including samples taken from shore. That said, if personnel time is at a greater premium in the spring or summer, winter surveying can be effective, particularly if sites are selected carefully and more sites are targeted with more samples and replication to offset lower detection probabilities.

This brings us to the question of how much sampling effort should be invested for a robust eDNA survey and how best to allocate that effort between sites, samples at sites and or qPCR replication. One starting place for this consideration is to use other published surveys as models of what has or has not been effective in various contexts. This is the approach I used to establish the surveys described in this study. There is no standardized sampling scheme for eDNA surveys in general and previous studies have varied widely in every level of sampling and testing. Some variation is related to hydrology or biology of the species. Sampling in rivers, for example, should follow a different plan than sampling in lakes or swamps (Pilliod et al., 2013; USFWS, 2013), and sampling sedentary resident species presents different spatiotemporal considerations than sampling migratory species. Since I sampled resident species in lakes, I concentrated on published studies of species inhabiting standing waters, although even within that group sampling has varied substantially. The number of sites sampled by various investigators ranges from as few as one to as many as 24 (Thomsen et al., 2012a; Takahara et al., 2013; Eichmiller et al., 2014; Lacoursière-Roussel et al., 2015). However, studies with a large number of sites often had a goal of mapping fish densities (Eichmiller et al., 2014; Ericson et al., 2016) or inferring abundances (Lacoursière-Roussel et al., 2015), rather than just species detection.

Since my goal was to confirm prior species reports, I hypothesized that fewer sites would be sufficient. Likewise, from MDIFW's perspective, the cost effectiveness and value of eDNA surveys, as an alternative to other survey methods to confirm exotic species reports, would be much less if it required large field sampling efforts. Notably, some of the water bodies I sampled were very small (0.2-0.6 km²), whereas others were very large (over 36 km²) and three samples may be less adequate for larger waterbodies than smaller ones, particularly during the early phases of population establishment. However, given that a major goal of my study was also to estimate detection probabilities from hierarchical occupancy models, I opted to set my site sampling effort at a fixed number of sites across all the waterbodies I tested.

As with numbers of sampling sites, there is considerable variation among published studies in the number of samples taken per site. Indeed, many studies have sampled only once per site in a water body (Takahara et al., 2013; Eichmiller et al., 2014; Erickson et al., 2016; Lacoursière-Roussel et al., 2015). However, single samples per site would be a risky prospect when surveying only three sites per waterbody. Moreover, I *a priori* suspected that eDNA might be very low in some sites with unconfirmed reports, was uncertain about the amount of eDNA that might be encountered in winter, and wanted to apply hierarchical occupancy models that depend on replication at all sampling levels, so I chose to sample thrice per site, following Hunter (2015) and Thomsen (2012a). Sample volumes are somewhat more standard across studies, with 1 L samples being the most common for studies using filtering of whole water (Takahara et al., 2013; Eichmiller et al., 2014; Hunter et al., 2015; Lacoursière-Roussel et al., 2015), although some studies have taken as little as 50 ml or 15 ml of water (Thomsen et al., 2012a ; Erickson et al., 2016).

Finally, at the level of number of qPCR replicates per field sample, I ran four replicates per water sample initially, and then ran another four samples for a given lake-species combination if no replicates were positive across all sites and samples. This iterative approach to qPCR replication reflects my study goal of confirming prior angler reports and represented a cost effective means to allocate lab resources where they were needed most, namely reports where the species and its eDNA are likely least abundant

and most difficult to amplify. For comparison, many of the qPCR studies that served as a model for my surveys ran three technical replicates per water sample (Takahara et al., 2012; Thomsen et al., 2012a; Eichmiller et al., 2014; Hunter et al., 2015), but some studies have run as many as twelve replicates, particularly those using standard PCR instead of qPCR (Herder et al., 2014; Jerde et al., 2011).

While prior studies can serve as models for initial survey design, and my surveys compare favorably in terms of replication, the most informative way to evaluate the sufficiency of a survey is to estimate its actual statistical power using the empirical data. In my study I used three-tiered hierarchical occupancy models to estimate detection probabilities at the three levels of sampling: the probability that eDNA is present at a site within a lake where the species occurs, that eDNA is collected in a sample at that site, and the probability that eDNA is detected in a qPCR replicate of positive sample. Many traditional field surveys and early eDNA surveys failed to consider imperfect detection (Gu & Swihart, 2004; Willoughby et al., 2016). Mordecai (2011) described the 3-tiered hierarchical Bayes occupancy model used to model not only base occupancy, but temporal occupancy (at the time of survey) and detection for the Louisiana waterthrush (*Parkesia motecilla*). Although this study relied on sight and sound observations of target birds, the same model is easily relatable to eDNA detection data. The first application of such a hierarchical model using eDNA was by Schmidt (2013) to reanalyzed data from a set of twenty ponds surveyed genetically for the chytrid fungus *Batrochochytrium dendrobatidis*. However, more analogous to my study, Hunter (2015) conducted an eDNA survey in southern Florida for the invasive Burmese Python (*Python molurus bivittatus*), and used these occupancy models to make estimate detection probabilities at each level of the sampling process. They concluded that although they often failed to find eDNA of their target species in a single sample, they nearly always found it in at least one of the three samples taken per site, suggesting that multiple samples greatly increased the probability of detection and that, in their case, was the level with the most influence on overall detection.

I employed an occupancy model framework very similar to Hunter (2015) and to Kery and Royle (2016) to estimate the above detection probabilities for each lake-species survey where I had at least one positive qPCR result. Because of the substantial number of lakes and multiple target species, I had a large

number of independent survey estimates of these probabilities compared to prior studies, and presumably even greater ability to capture the range of likely detection conditions than these prior studies. I modeled the seasons separately and found there was a gap in detection probability on all levels that might not be fully apparent from the overall patterns of lake-level detections. Specifically, mean detection probabilities tended to be overall lower in winter than spring. Indeed, examining my own field survey design against the cumulative probability curves estimated by my occupancy models, it appears that my spring sampling had very high power to detect my target species, whereas my winter sampling may not have. In spring, my mean probability of amplifying eDNA in at least one of four qPCR replicates was 98.7%, my mean probability of eDNA being present in at least one of three samples per site was 93.0%, and my mean probability that eDNA was present at a minimum of one of my three sites was 90.0%. By comparison, four qPCR replicates would on average only provide 86.0% probability of amplification in winter. Even with the eight qPCR replicates I employed for a subset of sites without initial detections, my detection probability only increased to 88.0%. At the sample level, three samples per site only provided a mean cumulative probability of about 92.2% for sites where eDNA concentrations had been high enough to employ just four replicates, and this dropped under 80.1% at sites where eight qPCR replicates proved necessary. Finally, at the effort level of three sites per lake, those surveys where 4 qPCR replicates had been adequate proved to have a cumulative amplification probability of 95.0%, suggesting three sites was on average suitable for detection. However, for those surveys requiring eight qPCR replicates, the mean cumulative probability for three sites was only 78.0% odds of detecting the target species. Hence, while detections were still generally in my favor for winter samples, there were non-insignificant odds that I would have missed the species in some sites. The fact that I increased my qPCR replicates to eight for sites without initial detections may have helped offset some of this loss of power, but the fact that some surveys were positive in spring/summer when they were not, or were questionable, in winter fits this outcome as does the larger number of sites per waterbody with positive detections in spring/summer than in winter.

With this in mind, failure to detect Black Crappie in Spectacle Pond, Largemouth Bass in Scraggly Pond and Smallmouth Bass in Kimball Pond during both winter and spring surveys could reflect false negatives due to limited power, but those failures could also represent real conditions in those waterbodies. Both Spectacle and Scraggly Ponds represent unconfirmed angler reports and such reports are prone to a reasonably high degree of uncertainty due to identification errors. In contrast, Kimball Pond is a location where MDIFW did previously confirm presence of Smallmouth Bass by capturing a single large adult fish after during spring trap-netting in 2013. No other bass were caught in subsequent years by MDIFW personnel, nor did they receive any further reports from the public. Given the small size of Kimball Pond, at just 0.22 km², MDIFW concluded that removal of that one fish had likely eradicated Smallmouth from the system. Consistent with this, the very small size of Kimball Pond makes it somewhat unlikely I would have missed sampling Smallmouth eDNA in two distinct seasons were it actually present. In this respect, my failure to detect Smallmouth Bass eDNA in Kimball Pond could be considered a confirmation of that eradication effort. Indeed, other studies have recently shown that eDNA can be used to confirm successful species eradications (Dunker et al., 2016).

Nonetheless what might a more optimal survey design look like based on the findings from my occupancy models. To answer that question, it may be informative to look at the range of detection probabilities when considering conservative survey designs. The logic for this is simply the high inferential priority often placed on having high confidence of detection under even suboptimal conditions. Hence, while the probabilities of a positive qPCR replicate were often much higher in spring than winter samples, the lower ranges of these estimated detection probabilities were actually very similar. Using these lower bounds, a conservative study design would include between seven to eight qPCR replicates. The same lower bound similarity was seen across seasons for the probabilities of presence of eDNA in a given sample at a site, with between about five and six samples providing greater than 95% probability of collecting eDNA. Finally, at the level of sites, the probability of encountering eDNA was similar for high detection probability sites, but ranged much lower for winter detections. This finding is consistent with fish, and their eDNA, being more patchily distributed and localized in lakes during winter. However,

estimates from the lowest detection probability survey in winter suggest that at least ten sites would be needed to have greater than 95% probability of sampling eDNA in that system. All that said, it should be kept in mind that these levels of replication are based on worst case estimates of detection probabilities from our data and might be excessive given that the lower bound estimates we obtained were often associated with much larger estimation error. A pragmatically better empirical recommendation for sampling might come from using the mean cumulative probability curve from just the subset of occupancy model fits from actual surveys of unconfirmed angler reports, as those are likely to be typical of the types of situations MDIFW will face in the future. Based on this subset of surveys, a recommendable design for >95% confidence at the qPCR replicate, sample and site levels would be 4, 3 and 4-5 in spring/summer, respectively, versus 7, 5 and 7 for winter.

Through this study I demonstrated the efficacy of my qPCR assays and winter sampling technique at detecting *Centrarchids* in Maine waters, including positively detecting them in five previously unknown locations. Although there are logistical challenges with winter sampling, including extra gear that requires full sterilization, potentially lower mean eDNA abundance with greater eDNA spatial heterogeneity, and limitations due to weather conditions, the technique also offers a viable alternative to summer/spring sampling that reduces potential risks of false positives from outside sources. My occupancy models have shown that, although I was successful at detecting targets in most waterbodies, and that my sampling had relatively high power to detect my target species in many of my surveys, particularly in spring, the design I employed could in principle be improved upon by modestly increasing the number of sites sampled and samples per site, and by using an adaptive qPCR replication approach that increases the number of replicates when initial assays are negative. These modifications are likely to be most beneficial for winter sampling.

Additional cost savings could come from applying further adaptive sampling and processing approaches. For example, while more water samples might be collected at more sites per lake, that does not imply they must all be analyzed at the same time. A subset could be run initially, with the remainder added if and when initial sample processing fails to detect a target. Again, such an approach would

effectively reallocate costly lab resources away from sites where detection is easy and toward sites where detection is more challenging. Another possible approach could even be to pool water samples across multiple sites within a waterbody for extraction or amplification to effectively sample more geographic space with fewer lab replicates. However, this pooling approach would come at a cost of knowing where target eDNA was sampled in a waterbody, and, unlike the above adaptive approach, my estimates of detection probabilities cannot be applied to assess the power of this pooling method.

Importantly, while I have field trialed one study design, and provided some recommendations on others, the results of the present study provide data that can be used to customize eDNA sampling to meet whatever detection criteria are requested by MDIFW or other stakeholders. In this way, managers can allocate their eDNA resources among surveying more sites, the same sites in more years, or fewer sites and years but with greater confidence per survey. This is important given that not all exotic introductions present the same risks due to differences in exotic species attributes, waterbody attributes, community vulnerabilities, risks of spread, and local management objectives. However, even when very high confidence of presence or absence is required, the total cost in personnel time, gear and assay expenses associated with eDNA are likely to be a fraction of the expenses of many traditional fisheries survey approaches. In addition, the sampling techniques, particularly in spring/summer, are simple to implement that with minimal training and might be conducted by non-scientist personnel such as lake associations, school groups or citizen scientists, further reducing the time burden of government personnel. eDNA is a powerful tool in exotic and rare species detection, and when used in conjunction with a biologically and statistically informed sampling scheme, can provide agencies, NGOs and citizens groups with unprecedented capacity to track and potentially mitigate the costs of emerging species invasions.

Future work will likely provide even more primer-probe toolsets, allowing for a wider picture of both non-native and native species presence within Maine and surrounding areas. Additionally, since the current toolsets utilize qPCR, there is the potential to relate estimated starting DNA copy numbers in water samples to target species abundance. This might be done at the scale of whole waterbodies or in a more localized fashion to map spatio-temporal dynamics of species within waterbodies (Takahara et al.,

2012; Pilliod et al., 2013; Lacoursière-Roussel et al., 2015; Erickson et al., 2016). Both of these approaches could further refine management of exotic and native species. However, the field design and analyses I explored in this study are very specific to sensitive presence versus absence detection of species and may not be optimal for these other objectives. Indeed, these abundance approaches will require additional data and surveys to calibrate, including independent estimates of relative or absolute abundances from traditional survey approaches (Lacoursière-Roussel et al., 2015). However, once established these eDNA tools might again provide substantial cost and time savings to allow monitoring in more waterways while preserving critical field resources for targeted management interventions that cannot be addressed with eDNA, such as those that deal with fish size, fish health, control of unwanted species and enhancement of valued fisheries.

REFERENCES

- Ahrenstorff, T. D., Sass, G. G., & Helmus, M. R.. The influence of littoral zone coarse woody habitat on home range size, spatial distribution, and feeding ecology of Largemouth Bass (*Micropterus salmoides*). *Hydrobiologia*, 2009;623(1), 223-233.
- Annett C. Hunting behavior in Florida Largemouth Bass, *Micropterus salmoides floridanis*, in a channelized river. *Environ Biol Fishes*. 1998; 53:75-87.
- Arslan T, Phelps RP. Production of monosex male Black Crappie, *Pomoxis nigromaculatus*, populations by multiple androgen immersion. *Aquaculture*. 2004;234(1-4):561-573.
- Bonney F. Brook Trout management plan. *Maine Department of Inland Fisheries and Wildlife*. 2009.
- Boucher D. Landlocked Salmon management plan. *Maine Department of Inland Fisheries and Wildlife*. 2004.
- Brewer SK, Rabeni CF. Interactions between natural-occurring landscape conditions and land use influencing the abundance of riverine Smallmouth Bass, *Micropterus dolomieu*. *Can J Fish Aquat Sci*. 2011;68(11):1922-1933.
- Dauwalter DC, Fisher WL. University of Notre Dame spawning chronology, nest site selection and nest success of Smallmouth Bass during benign streamflow conditions. *Am Midl Nat*. 2015;158(1):60-78.
- Deiner K, Altermatt F. Transport distance of invertebrate environmental DNA in a natural river. *PLoS One*. 2014;9(2).
- Dejean T, Valentini A, Duparc A. Persistence of environmental DNA in freshwater ecosystems. *PLoS One*. 2011;6(8):4-8.
- Dunker KJ, Sepulveda AJ, Massengill RL, et al. Potential of Environmental DNA to Evaluate Northern Pike (*Esox lucius*) Eradication Efforts: An Experimental Test and Case Study. *PLoS One*. 2016;11(9).
- Eichmiller JJ, Bajer PG, Sorensen PW. The relationship between the distribution of common carp and their environmental DNA in a small lake. *PLoS One*. 2014;9(11).
- Ellender BR, Weyl OLFF, Swartz ER. Invasion of a headwater stream by non-native fishes in the Swartkops River system, South Africa. *African Zool*. 2011;46(1):39-46.
- Ficetola GF, Pansu J, Bonin A, et al. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Mol Ecol Resour*. 2015;15(3):543-556.
- Erickson RA, Rees CB, Coulter AA, et al. Detecting the movement and spawning activity of bigheaded carps with environmental DNA. *Mol Ecol Resour*. 2016;16(4):957-965.
- Frost F, Arctic Charr management plan. *Maine Department of Inland Fisheries and Wildlife*. 2001.
- Gu W, Swihart RK. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biol Conserv*. 2004;116(2):195-203.
- Hanson KC, Cooke SJ, Suski CD. Assessment of Largemouth Bass (*Micropterus salmoides*) behaviour and activity at multiple spatial and temporal scales utilizing a whole-lake telemetry array. *Hydrobiologia*. 2007;582(1):243-256.

- Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. *Genome Methods*. 1996;6:986-994.
- Herder J, Valentini A, Bellemain E, Environmental DNA a review of the possible applications for the detection of (invasive) species. *Stiching RAVON, Nijmegen*. 2014;(November):Report 2013-2104.
- Huchzermeyer C, Weyl O, Cowley P. Evaluation of acoustic transmitter implantation and determination of post-translocation behaviour of Largemouth Bass *Micropterus salmoides* in a South African impoundment. *African J Aquat Sci*. 2013;38(2):229-236.
- Hunter ME, Oyler-McCance SJ, Dorazio RM, et al. Environmental DNA (eDNA) sampling improves occurrence and detection estimates of invasive Burmese pythons. *PLoS One*. 2015;10(4):1-17.
- Jane SF, Wilcox TM, Mckelvey KS, et al. Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Mol Ecol Resour*. 2015;15(1):216-227.
- Jerde CL, Mahon AR, Chadderton WL, Lodge DM. "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conserv Lett*. 2011;4(2):150-157.
- Johnson P, Lake Trout management plan. *Maine Department of Inland Fisheries and Wildlife*. 2001.
- Jordan R, Black Bass management plan. *Maine Department of Inland Fisheries and Wildlife*. 2001.
- Josefsen MH, Löfström C, Sommer HM, Hoorfar J. Diagnostic PCR: Comparative sensitivity of four probe chemistries. *Mol Cell Probes*. 2009;23(3-4):201-203.
- Kéry, M, Royle, JA. *Applied Hierarchical Modeling in Ecology*. Elsevier:2016.
- Kirillin G, Leppäranta M, Terzhevik A, Granin N, Bernhardt J, Engelhardt C, Efremova T, Golosov S, Palshin N, Sherstyankin P, Zdorovenova G, Zdorovenov R. Physics of seasonally ice-covered lakes: A review. *Aquat Sci*. 2012;74(4):659-682.
- Kutyavin I V, Afonina IA, Mills A, et al. 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res*. 2000;28(2):655-661.
- Kwok S, Kellogg DE, Mckinney N, et al. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nucleic Acids Res*. 1990;18(4):999-1005.
- Lacoursière-Roussel A, Côté G, Leclerc V, Bernatchez L, Cadotte M. Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. *J Appl Ecol*. 2015;53(4):1148-1157.
- Lasenby, T. A., & Kerr, S. J. Bass transfers and stocking: An annotated bibliography and literature review. *Fish and Wildlife Branch, Ontario Ministry of Natural Resources. Peterborough, Ontario*. 2000
- Livak KJ, Flood SJ, Marmaro J, et al. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *Genome Res*. 1995;4:357-362.
- Lodge DM, Williams S, Macisaac HJ, et al. Biological Invasions: Recommendations for U.S. policy and management. *Ecol Appl*. 2006;16(6):2035-2054.
- Lucas, J. Minor sportfish management plan. *Maine Department of Inland Fisheries and Wildlife*. 2002.
- Mackenzie DI, Nichols JD, Lachman GB, et al. Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. *Ecology*. 2002;83(8):2248-2255.
- Magnuson, J. J., Benson, B. J., & McLain, A. S. Insights on species richness and turnover from long-term ecological research: fishes in north temperate lakes. *American Zoologist*, 1994; 34(3), 437-451.

- Mesing, C. L., & Wicker, A. M. Home range, spawning migrations, and homing of radio-tagged Florida Largemouth Bass in two central Florida lakes. *Transactions of the American Fisheries Society*, 1986; 115(2), 286-295.
- Mikeska T, Dobrovic A. Validation of a primer optimisation matrix to improve the performance of reverse transcription - quantitative real-time PCR assays. *BMC Res Notes*. 2009; 2:112.
- Mills LS, Pilgrim KL, Schwartz MK, McKelvey K. Identifying lynx and other North American felids based on MtDNA analysis. *Conserv Genet*. 2000;1:285-288.
- Minamoto T, Yamanaka H, Takahara T, Honjo MN, Kawabata Z. Surveillance of fish species composition using environmental DNA. *Limnology*. 2012;13(2):193-197.
- Minns CK. Allometry of home range size in lake and river fishes. *Can J Fish Aquat Sci*. 1995;1508(52):1499-1508.
- Mordecai RS, Mattsson BJ, Tzilkowski CJ, Cooper RJ. Addressing challenges when studying mobile or episodic species: Hierarchical Bayes estimation of occupancy and use. *J Appl Ecol*. 2011;48(1):56-66.
- Peat TB, Gutowsky LFG, Doka SE, et al. Comparative thermal biology and depth distribution of largemouth urban harbour of the Laurentian Great Lakes. *Can J Zool*. 2016;776(October):767-776.
- Phelps QE, Lohmeyer AM, Wahl NC, Zeigler JM, Whitley GW. Habitat characteristics of Black Crappie nest sites in an Illinois impoundment. *North Am J Fish Manag*. 2009;29(1):189-195.
- Piggott MP. Evaluating the effects of laboratory protocols on eDNA detection probability for an endangered freshwater fish. *Ecol Evol*. 2016;6(9):2739-2750.
- Pimentel D, Zuniga R, Morrison D. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol Econ*. 2005;52(3):273-288.
- Pilliod DS, Goldberg CS, Arkle RS, Waits LP, Richardson J. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Can J Fish Aquat Sci*. 2013;70(8):1123-1130.
- Pope K, Willis D. Environmental characteristics of Black Crappie (*Pomoxis nigromaculatus*) nesting sites in two South Dakota waters. *Ecol Freshw Fish*. 1997; 6:183-189.
- Reed JR, Pereira DL. Relationships between shoreline development and nest site selection by Black Crappie and Largemouth Bass. *North Am J Fish Manag*. 2009;29(4):943-948.
- Ridgway M, Shuter B. Effects of displacement on the seasonal movements and home range characteristics of Smallmouth Bass in Lake Opeongo. *North Am J Fish Manag*. 1996;16(November):371-377.
- Royle JA, Link WA. Generalized Site Occupancy Models Allowing for False Positive and False Negative Errors. *Ecology*. 2006;87(4):835-841.
- Sevilla RG, Diez A, Norén M, et al. Primers and polymerase chain reaction conditions for DNA barcoding teleost fish based on the mitochondrial cytochrome b and nuclear rhodopsin genes. *Mol Ecol Notes*. 2007;7(5):730-734.
- Shokralla S, Spall JL, Gibson JF, Hajibabaei M. Next-generation sequencing technologies for environmental DNA research. *Mol Ecol*. 2012;21(8):1794-1805.
- Schmidt BR, Kéry M, Ursenbacher S, Hyman OJ, Collins JP. Site occupancy models in the analysis of environmental DNA presence/absence surveys: A case study of an emerging amphibian pathogen. *Methods Ecol Evol*. 2013;4(7):646-653.

Shuter BJ, Finstad AG, Helland IP, Zweimüller I, Hölker F. The role of winter phenology in shaping the ecology of freshwater fish and their sensitivities to climate change. *Aquat Sci.* 2012;74(4):637-657.

Stadhouders R, Pas SD, Anber J, Voermans J, Mes THM, Schutten M. The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5' nuclease assay. *J Mol Diagnostics.* 2010;12(1):109-117.

Strayer DL. Alien species in fresh waters: ecological effects, interactions with other stressors, and prospects for the future. *Freshw Biol.* 2010; 55:152-174.

Takahara T, Minamoto T, Yamanaka H, Doi H, Kawabata Z. Estimation of fish biomass using environmental DNA. *PLoS One.* 2012;7(4)

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

Taylor S, Wakem M, Dijkman G, Alsarraj M, Nguyen M. A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines. *Methods.* 2010;50(4): S1-S5.

Thomsen PF, Kielgast J, Iversen LL,. Monitoring endangered freshwater biodiversity using environmental DNA. *Mol Ecol.* 2012a; 21(11):2565-2573.

Thomsen PF, Kielgast J, Iversen LL, Møller PR, Rasmussen M, Willerslev E. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS One.* 2012b;7(8)

Turner CR, Uy KL, Everhart RC. Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biol Conserv.* 2014;183:93-102.

U.S. Fish and Wildlife Service (2013) Quality assurance project plan (QAPP): eDNA monitoring of bighead and silver carps. 89 pp.

Vander Zanden MJ, Olden J, Thorne J, Mandrak N. Predicting occurrences and impacts of Smallmouth Bass introductions in north temperate lakes. *Ecol Appl.* 2004;14(1):132-148.

Vander Zanden MJ, Hansen GJA, Higgins SN, Kornis MS. A pound of prevention, plus a pound of cure: Early detection and eradication of invasive species in the Laurentian Great Lakes. *J Great Lakes Res.* 2010;36(1):199-205.

Wilcox TM, McKelvey KS, Young MK, et al. Robust Detection of Rare Species Using Environmental DNA: The Importance of Primer Specificity. *PLoS One.* 2013;8(3).

Wilcox TM, McKelvey KS, Young MK, et al. Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char *Salvelinus fontinalis*. *Biol Conserv.* 2016;194:209-216.

Willoughby JR, Wijayawardena BK, Sundaram M, Swihart RK, DeWoody JA. The importance of including imperfect detection models in eDNA experimental design. *Mol Ecol Resour.* 2016;16(4):837-844.

Willerslev E, Hansen AJ, Binladen J,. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science.* 2003; 300(5620):791-795.

Willerslev E, Cooper A. Ancient DNA. *Proc R Soc B Biol Sci.* 2005;272(December 2003):3-16.

Yonekura R, Kita M, Masahide Y. Species diversity in native fish community in Japan: comparison between non-invaded and invaded ponds by exotic fish. *Ichthyol Res.* 2004;51(2):176-179.

Zaret TM, Paine RT. Species introduction in a tropical lake: A newly introduced piscivore can produce population changes in a wide range of trophic levels. *Sci [Science (Wash)]*. 1973;182(4111):449-455.

APPENDICES

Appendix A:

Table A.1: Alignments of target species primer-probe vs. related species

| Assay | Forward Primer (5'-3') | MGB Probe (5'-3') | Reverse Primer (5'-3') |
|------------|---------------------------|--------------------------------------|-------------------------------|
| BC | GCCTCTGCTTGCCACCCAAAT | GCAACTGCCTTCTCCTCCGTAGCA | CCGCAACATTCATGCCAATCGG |
| LMB | GCCTCTGC TGGC ACCCA GAT | GCAAC CGCTTCTC ATCCGT TGC | TCG AACATTCATGC TAAACGG |
| SMB | GCCTCTGC TGGC AC TCAAAT | GCAAC CGCTTCTC ATCCGT TGC | CCG AACATTCATGC TAAAT TGG |
| RBS | GC TATGC CT TGC CACCCAAAT | GCAAC CGCT TCTCT TCA GT TCGCA | CCGCAACATTCAC GCCAA CCGG |
| PSS | GCCTATGC CT AGCCACCCAAAT | GCAAC CGCT TCTCTCT AGT TCGCA | TCG TAAACATTCAC GCCAA CCGG |
| BG | GCCTCTGC CT AGCA ACCCAAAT | GCAAC CGCTTCTCT TCA GTAGCC | CCG TAAACAT TCA CAGGCCAA CCGG |
| LMB | CGCTGCCGCCACAGTAATCCAT | CCCCCTGGGACTAAACTCTGACGC CG | TCGCAGCTCTCCTCATTGCCCT |
| BC | CGCTGCCG T CACAGT CAT CAT | TCCCCTGGG CTAAAT TCA GACGC GG | T TGC G TCTCCTCATTGCCCT |
| SMB | CGCTGC T GCCACAGTAAT TCA | CCCCCTAGGAT TGA ACTCTGACGC TG | TCGCAGC CTCTCATTGCCCT |
| RBS | TGCCGCCGCCAC T GTAAT TCA | TCCCCTGGG CTAAACTC TGA TGC AG | T TGCAGC CTCTCT TATTGCCCT |
| PSS | TGC AGCCGCCAC T GTAAT TCA | CCCATTAGG CTAAACTC TGA GACGC AG | T TGCAGC CTCTCTCATTGCCCT |
| BG | TGCCGCCGC AAC T GTAATCCA | CCCCCTAGG CT TAAACTC TAAACGC AG | T TGTAGC ACTCTCATTGCCCT |
| SMB | GGGTGTCTCCTCCATCCTAGGGGCC | TCAGACACCCTGTTTGTGGTCC GTCTT | TTATCGCTCCCAGTCCCTCGCTGC T |
| BC | TGGAGTCTCTCCATCTGGGGCT | TCAGACGCC TTTATTTGTCTGATCC GTCTT | CTTTC ACTCCAGTCTTGCTGC G |
| LMB | TGGTGTCTCCTCTATCTAGGGGCA | GCAACACACCCTTTTGTCTGATC CGTCTT | CTGTC ACTCCAGTCCCTCGCTGC G |
| RBS | TGGGTCTCTTCAATCTAGGGGCT | GCAACACACCCTGTTTGTCTGATC AGTCTT | CTTTC CTCCC GTCTTGCTGC A |
| PSS | TGGGTCTCTTCAATCTAGGGGCA | CCAGACACCCTGTTTGTCTGGTC CGTATT | CTCTC CTCCCAGTCTTGCTGC A |
| BG | TGGGTCTCTTCAATCCTGGAGCT | CCAGACCC TTTATTTGTCTGATCA GTCTT | CTTTC CTCCAGTCTTGCTGC A |
| SUN | CTCTACAAGGACCTCCTCGGCTTT | GCCCTCCTATTGCCCTGACTTCCC TGGCAC | TGCTCGGGGACCCAGACAACCTT |
| BC | TCTCTACAAGGACCTCTCGGTTT | G TCTCCTCTATTGCCCTTACTTCCC TAGCAT | TCTTAGGAGACCCGACAACCTT |
| LMB | CTCTATAAAGACTTGGTTT | GCTCTCCTCTATTGCCCTTACTTCA TAGCCT | T TCTAGGGGACCCGACAACCTT |
| SMB | CTCTACAAAGACCTCTTGGATT | GCCCTCCTCTATTGCCCTTACTTCA TAGCCT | TCTTAGGAGACCAGACAACCTT |
| BG | CTCTATAAAGACCTCTTGGTTT | GCTCTCTCTATTGCCCTTACTTCT TAGCAT | TGCTAGGAGACCAGACAACCTT |

Appendix B: Winter sampler effect

Comparison of spring results with sampler vs. without sampler using paired two sample t-test showed no significant differences in C_q value ($p = 0.4637$) or number of positive amplifications ($p = 0.7539$) between use or non-use of sampling device.

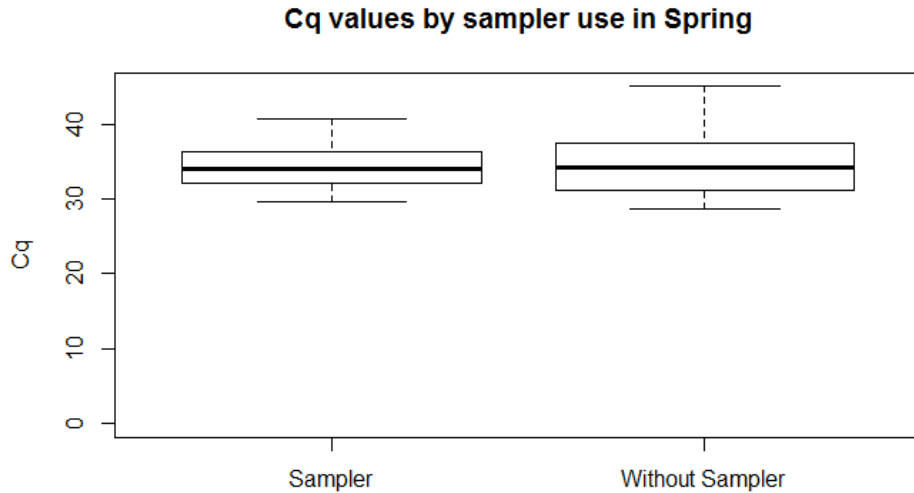


Figure B.1: C_q values by sampler use in spring. Whiskers represent full range excluding outliers. Box represents 75th and 25th quartile, horizontal band represents 50th quartile.

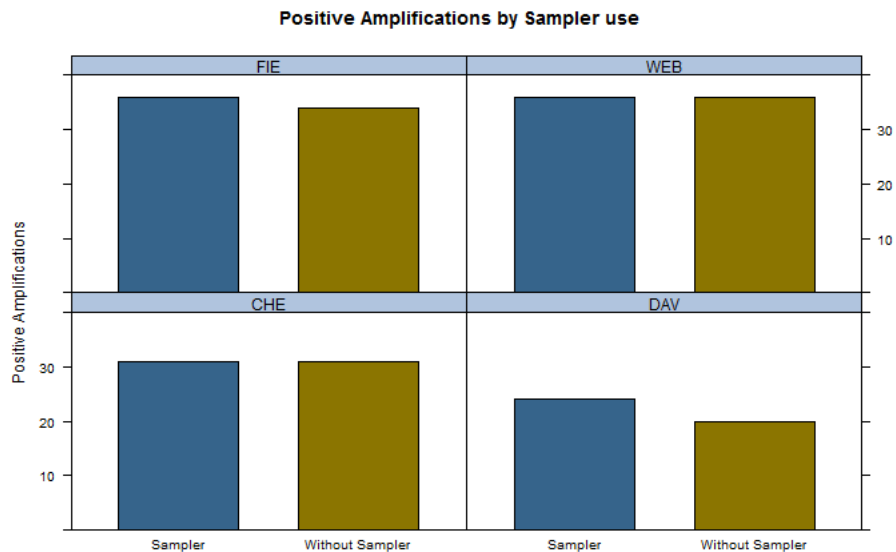


Figure B.2: Positive amplification by sampler use. Total number of target species DNA amplifications by lake and with/without sampler use.

Appendix C: R code for hierarchical occupancy models

R code for 3-tiered hierarchical model for WinBUGS.

```
#load required packages
library(AHMbook)
library(car)
library(R2WinBUGS)
#import data for lake 1
pcr1 <- read.csv("C:/occu/spring/BOG_LS_pcr1.csv", header=FALSE)
pcr2 <- read.csv("C:/occu/spring/BOG_LS_pcr2.csv", header=FALSE)
pcr3 <- read.csv("C:/occu/spring/BOG_LS_pcr3.csv", header=FALSE)
pcr4 <- read.csv("C:/occu/spring/BOG_LS_pcr4.csv", header=FALSE)

A <- array(as.numeric(NA), dim = c(3,3,4)) #create empty array
A[,,1]<-as.matrix(pcr1) #data into array
A[,,2]<-as.matrix(pcr2)
A[,,3]<-as.matrix(pcr3)
A[,,4]<-as.matrix(pcr4)
y <- A
str( win.data <- list(y = y, n.pond = dim(y)[1], n.samples = dim(y)[2], n.pcr = dim(y)[3] ))

# Define model in BUGS language
###Without Covariates

sink("model.txt")
cat("
  model {
    # Priors and model for params
    int.psi ~ dunif(0,1)    # Intercept of occupancy probability (sites)
    for(t in 1:n.samples){
      int.theta[t] ~ dunif(0,1) # Intercepts availability probability (samples)
    }
  }
"
```



```

for(t in 1:n.pcr){
int.p[t] ~ dunif(0,1) # Intercepts detection probability (1-PCR error)
}

# 'Likelihood' (or basic model structure)
for (i in 1:n.site){
# Occurrence in site i
z[i] ~ dbern(psi[i])
logit(psi[i]) <- logit(int.psi)
for (j in 1:n.samples){
# Occurrence in sample j
a[i,j] ~ dbern(mu.a[i,j])
mu.a[i,j] <- z[i] * theta[i,j]
logit(theta[i,j]) <- logit(int.theta[j])
for (k in 1:n.pcr){
# PCR detection error process in sample k
y[i,j,k] ~ dbern(mu.y[i,j,k])
mu.y[i,j,k] <- a[i,j] * p[i,j,k]
logit(p[i,j,k]) <- logit(int.p[k])
}
}
tmp[i] <- step(sum(a[i,])-0.1)
}

# Derived quantities
sum.z <- sum(z[]) # Total # of occupied sites in lake
sum.a <- sum(tmp[]) # Total # of samples with presence
m.pcr<- mean(int.p[]) #mean p (across qPCR replicates)
m.sample<- mean(int.theta[]) # mean theta (across samples)
} # end model
",fill=TRUE)

```

```

sink()
# Initial values
zst <- apply(y, 1, max)    # inits for presence (z)
ast <- apply(y, c(1,2), max) # inits for availability (a)
inits <- function() list(z = zst, a = ast, int.psi = 0.5)
# parameters
params <- c("int.psi", "int.theta", "int.p", "sum.z", "sum.a", "m.pcr", "m.sample")
# MCMC setting
ni <- 5000 ; nt <- 2 ; nb <- 1000 ; nc <- 3
# Call WinBUGS and summarize posterior
out <- bugs(win.data, inits, params, "model.txt", n.chains = nc, n.thin = nt, n.iter = ni, n.burnin = nb,
debug = TRUE, bugs.dir = "c:/Program Files (x86)/WinBUGS14") #####bugs.dir="Location of
WinBUGS" ← put location of winbugs
print(out, 3)

```

BIOGRAPHY OF THE AUTHOR

Geneva was born in Portland, Maine. She didn't graduate from high school. She earned a B.S. in Biology from UMaine in 2013. She worked in California for the National Park Service before returning to school. She is a candidate for a Master of Science degree in Marine Biology from the University of Maine in December, 2016.