Molecular techniques for the identification of freshwater fish species for environmental monitoring programs

by

Emily Hulley

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science (MSc) in Biology

The Faculty of Graduate Studies
Laurentian University
Sudbury, Ontario, Canada

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis Titre de la thèse	Molecular techniques for the identification of freshwater fish species for environmental monitoring programs
Name of Candidate Nom du candidat	Hulley, Emily
Degree Diplôme	Master of Science
Department/Program Département/Programme	Biology
Date of Defence Date de la soutenance	October 26, 2018

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Douglas Boreham
(Supervisor/Directeur de thèse)

Dr. John Gunn
(Committee member/Membre du comité)

Dr. Kabwe Nkongolo
(Committee member/Membre du comité)

Dr. Gustavo Ybazeta
(External Examiner/Examinateur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. David Lesbarrères
Monsieur David Lesbarrères
Dean, Faculty of Graduate Studies
Doyen, Faculté des études supérieures

Abstract

Reliable species identification methods are important for industrial environmental monitoring programs. Probe based real-time PCR (qPCR) provides an accurate, cost-effective and high-throughput method for species identification. Here we present the development and validation of species-specific primers and probe for the identification of eight freshwater fish species. The development of a fully automated species-decoder algorithm allowed for target species identification with 100% accuracy while completely removing any false-positive detection of non-target species. Furthermore, the probe-based qPCR technique utilized in this study is substantially more cost-effective and time efficient than DNA barcoding and morphological identification methods. The qPCR assays were also highly sensitive and accurately detected target species from collected environmental DNA (eDNA) samples. In summary, probe-based multiplex qPCR assays provide a rapid and accurate method for freshwater fish species identification and the methodology established in this study can be utilized for various other species identification initiatives.

Keywords

Freshwater fish species, environmental monitoring, cytochrome oxidase I, DNA barcoding, morphological identification, TaqMan real-time PCR, multiplex PCR, environmental DNA, next generation sequencing, species identification, species detection

Co-Authorship Statement

This thesis was written with myself as the primary author, however, several coauthors were directly involved in several aspects of the research.

Appendix A includes the primary co-authorship of "DNA barcoding vs. morphological identification of larval fish and embryos in Lake Huron: advantages to a molecular approach" between Emily Hulley and Natalie Taylor. Natalie Taylor conducted the original research for DNA barcoding the larval fish and embryos and prepared the original manuscript. Emily Hulley has been the corresponding author for publication with the Journal of Great Lakes Research. She has also been primarily responsible for the major revisions to the manuscript and tables from reviewers, associate editors and collaborating authors.

Acknowledgments

There are countless people I would like to thank for their help and support during my thesis. First and foremost, I would like to thank my supervisor Dr. Douglas Boreham for the opportunities and support you provided to me throughout my degree. I would also like to thank my Boreham lab mates for an amazing two years. Additionally, I would like to say a special thank you to Dr. Sujeenthar Tharmalingam and Dr. Chris Thome, who were always there, no matter how busy, to provide me with guidance and assistance whenever I needed it.

I would also like to thank my family and friends for their endless love, support and encouragement. If it hadn't been for my incredible parents and family, I would not be who or where I am today. I am so fortunate to have such an amazing and strong family for role models.

Table of Contents

List of Tables

List of Figures

List of Acronyms

BLAST	Basic Local Alignment Search Tool
BOLD	Barcode of Life Database
Bp	Base pairs
CANDU	Canadian Deuterium Uranium
COI	Cytochrome oxidase I
Ct	Cycle threshold
DNA	Deoxyribonucleic Acid
eDNA	Environmental DNA
NCBI	National Center for Biotechnology Information
NGS	Next Generation sequencing
PCR	Polymerase Chain Reaction
SS	Sanger sequencing
qPCR	Real-time PCR

List of Appendices

Chapter 1: Introduction

Many industrial processes utilize water for cooling as part of their daily operations. Once-through cooling systems are commonly used by largescale industries such as nuclear power plants, petrochemical plants, petroleum refineries, paper mills, natural gas and food processing plants. Once-through cooling systems intake a large volume from a nearby source (rivers, lakes, the ocean) where it absorbs heat from the condensers as its circulates through the plant before being discharged to the source as a warmed surface effluent (1).

Bruce Power is a nuclear power plant located on the eastern shores of Lake Huron (Tiverton, Ontario). The power plant consists of two nuclear generating stations, Bruce A and Bruce B, each of which house four Canada Deuterium Uranium (CANDU) reactors (2). These eight units are all currently operational and produce up to 6,400 megawatts of electricity, accounting for over 30% of Ontario's electricity (2). Bruce Power uses two offshore intakes that draw water from Lake Huron into their once-through cooling systems, one for each station. Bruce A processes up to 193,000L/s and Bruce B 170,000L/s of water to cool their reactors. The system utilizes a series of grates and velocity caps to prevent surrounding wildlife from being entrained (i.e., a fish is drawn into a water intake and cannot escape) or impinged becoming impinged (i.e., a fish is held in contact with the intake screen and is cannot free itself). These caps reduce the water current to approximately 15cm/s, which adult and juvenile fish can outswim and avoid becoming entrained (2). However, small fish such as embryo, larval and some juveniles have limited to no swimming capabilities and as a result become impinged during water intake (2). The number of and types of species that become impinged or entrained varies greatly from year to year. In 2013, almost 25,000 individuals were impinged with yellow perch and rainbow smelt as the top species impinged at 7362 and 5573 respectively (2). The number of impinged individuals was significantly lower in 2014 at just over 16,000, with gizzard shad and yellow perch at the top species impinged at 9525 and 2207 respectively (2).

Morphological based identification has traditionally been the method used by environmental monitoring programs to identify impinged fish species at its intake points in once-through cooling systems. This technique utilizes key external morphological features, such as the jaw, scales, colour, fins etc., to differentiate and identify specimens to their species-level. However, this technique has proven to have several limitations, which is often compounded when individuals are impinged because they can be damaged or missing key identification traits. Incorrect identification can also arise with cryptic variation or intraspecific phenotypic plasticity of a trait (3–7). As well, key differentiating characteristics for morphological identification may only be effective for a particular life stage or gender. This issue often arises during the identification of larval fish as individuals can rapidly change during development and appear quite different from individuals that are from the same species (3–6). Additionally, many larval fish species share extremely similar morphology to those from another species (3–6). A study done by Ko *et al.* 2013 compared the accuracy of five different laboratories morphological identification techniques of larval fishes and found the accuracy rates between them to be extremely low; 80.1% for family-level, 41.1% for the genus-level and 13.5% at the species-level. Therefore, morphological identification of larval fishes requires significant level of expertise and even then, it is still recommended that identification be conservative and restricted to the family-level (4). Consequently, environmental monitoring programs require a significant degree of specificity and accuracy, requiring identification to the genus and species level. Inaccurate assessment of impacted individuals could result in misused conservation efforts on a less impacted species and subsequently under protection efforts for a highly impacted species.

Over the last decade, species identification using a molecular approach has been as an attractive alternative to morphological identification. In 2003, Hebert *et al.* from the University of Guelph proposed that a 650-base pair (bp) segment of the 5’ region of the mitochondrial cytochrome oxidase I (COI) gene could serve as a universal barcode for the animal kingdom (3,5,7–9). DNA barcoding identifies individuals to their respective species through the use of molecular tags which are based on short

fragment of DNA from a standardized region of a gene (COI, 16S ribosomal RNA, *cytochrome b* etc) (3,8). The DNA barcode of a species could allow for a standardized approach for the quick, efficient, robust and accurate species identification (8). DNA barcoding has been shown to be 99-100% accurate in identifying species when a comprehensive database for comparison is available. Specifically, with respect to freshwater fishes, databases such as Barcode of Life Database (BOLD) and GenBank have extensive publically available sequences (10). DNA barcoding has proven to be beneficial identification technique for forensic science, food traceability, cryptic communities and monitoring of invasive species (3,7,8,11,12). However, DNA barcoding does have some drawbacks. The first is that it requires several post-PCR processing steps, such as ethidium bromide staining, gel electrophoresis, PCR purification etc, which increases the cost and time to process, as well as the chance for product contamination. In addition, DNA barcoding relies on the variation within the mitochondrial DNA between species to accurately distinguish individuals. This is problematic for certain species, such as those within the genus *Coregonus*, which have a recent evolutionary divergence and have very little interspecies genetic variation and therefore cannot be accurately sequenced to the species level (7,13). Lastly, DNA barcoding does not allow for the detection of hybrid species as the mitochondrial genome is transferred through maternal inheritance, for example the species splake, a hybrid between a female lake trout (*Salvelinus namaycush*) and a male brook trout (*Salvelinus fontinalis*), would identify as lake trout through DNA barcoding (14,15).

Real-Time PCR (qPCR) coupled with Taqman probes provides an accurate and high-throughput alternative for species identification when compared to DNA barcoding and morphological identification (9,16–19). The TaqMan probe is dual-labeled with a fluorescent dye and a light-absorbing quencher. When the dye and quencher are in close proximity the probe is unable to emit a fluorescent signal. During PCR amplification the probe is hydrolyzed by *Taq* DNA polymerase during the extension phase, which results in the separation of the dyes and an increase in reporter fluorescence. Following repeated PCR cycles, intensity of the fluorescent signal is further increased due to the exponential amplification of the

PCR product (17,18,20). When primer and probes are designed to be species-specific, the fluorescent signal only occurs when the forward and reverse primers and probe detect their complementary sequence (17,20). Post-PCR processing steps are no longer required because the amplification is detected by the fluorescent probe (16–18). Compared to DNA barcoding, this not only increases efficiency, as results are obtained in a couple of hours, but also decreases the chance of contamination (17). In addition, TaqMan probes allow for multiple samples to be processed in a single reaction (multiplex), furthering increasing efficiency. Here, the probes can be linked to several fluorescent dyes that have distinct excitation-emission spectra. Multiplex capable qPCR machines can distinguish simultaneous emission readings and allow for multiple samples to be processed in a single well (9,16,17). Lastly, qPCR assays have the ability to accurately detect their target species when sample DNA is degraded or damaged. DNA barcoding requires high quality DNA to achieve successful amplification of the 650bp barcoding region of the COI gene to accurately identify samples. qPCR primer-probe sets only require 70 to 200bp regions of a sequence to detect their target, thus allowing greater chance of identification with degraded samples. TaqMan probe-based qPCR has been successfully developed for a variety of different applications including the detection of pathogen and bacterial strains (17,18), fish identification (16,20,21), food traceability (9,19,22) and most recently in the detection of species from environmental DNA (eDNA) samples (23–26).

eDNA has recently emerged as a novel species identification technique that is non-invasive, highly sensitive and cost effective (23,25,27,28). eDNA are short, species-specific fragments of DNA that are released from organisms into the environment through urination, defecation, epidermal cell shedding, secretion of gametes or mucus and carcasses (24,25,28–30). Depending on conditions, eDNA can remain in an aquatic environment for approximately 7 to 21 days (24). eDNA is collected from water samples through either precipitation or filtration onto micrometer pore-sized filters (31,32). In the literature, target species' eDNA has been detected through a variety of different methods including standard PCR, qPCR, digital drop PCR and next generation sequencing (NGS). eDNA has been used for the detection and

monitoring of aquatic invasive (23,31,33–35) and endangered species (27,36), as well as to estimate species biomass (25,26,28).

Here, we developed and validated species-specific probe-based qPCR assays for the cost-effective and high-throughput identification of eight freshwater fish species for industrial environmental monitoring programs in the Great Lake region, specifically Bruce Power. The eight species of interest included: yellow perch (*Perca flavescens*), lake whitefish (*Coregonus clupeaformis*), rainbow smelt (*Osmerus mordax*), smallmouth bass (*Micropterus dolomieu*), spottail shiner (*Notropis hudsonius*), brook trout, round whitefish (*Prosopium cylindraceum*) and deepwater sculpin (*Myoxocephalus thompsonii*). These species are not only ecologically important to their surrounding ecosystem, but many are economically important for commercial and recreational fisheries. Round whitefish and deepwater sculpin are not economically important but were included in the study as species of special concern because deepwater sculpin are listed as an endangered species (37) and round whitefish have been identified as potentially having an increased sensitivity to once-through cooling operations (38). Species-specific TaqMan probes were developed for the eight species of interest within the COI gene. A species decoder algorithm was also developed based on the qPCR results to automatically identify species to allow full automation of results analysis. Blinded experiments demonstrated that the combination of the species-specific assays and the automated species decoder resulted in 100% accuracy in detecting target species, with no false detection of non-target species. The primer-probe sets were also highly sensitive, detecting to DNA concentrations as low as 1 ng, which was adequate for positive species identification. We also conducted a cost and time comparison analysis between the qPCR assays designed in this study compared to morphological identification and DNA barcoding.

Next, we were interested in determining if the validated qPCR assays were able to detect their target species from eDNA samples. We selected Lake Laurentian (Sudbury, ON) as our study site as it is a popular location for recreational fishermen and contains yellow perch, one of our target species (39). Three 1L water samples were collected from six sites across the lake in March 2018. The yellow perch

qPCR assay detected yellow perch from five of the six sites, which was confirmed using Sanger sequencing and NGS. In addition, trends in the NGS results for yellow perch paralleled those obtained from the qPCR reactions.

Chapter 2: Development and Validation of Probe-Based Multiplex Real-Time PCR Assays for the Rapid and Accurate Detection of Fish Species

Development and Validation of Probe-Based Multiplex Real-Time PCR Assays for the Rapid and Accurate Detection of Fish Species.

(Original Research)

Emily N. Hulley, Sujeenthar Tharmalingam, Andrew Zarnke, Douglas R. Boreham

(In review for publication in PLOS One)

Abstract

Reliable species identification methods are important for industrial environmental monitoring programs. Probe based real-time polymerase chain reaction (qPCR) provides an accurate, cost-effective and high-throughput method for species identification. Here we present the development and validation of species-specific primers and probes for the cytochrome c oxidase (COI) gene for the identification of eight freshwater fish species. In order to identify novel primer-probe sets with maximum species-specificity, two separate primer-probe design criteria were employed. Highest ranked primer-probe sets from both methods were assayed to identify sequences that demonstrated highest specificity. Specificity was determined using control species from same genus and non-target species from different genus. Selected primer-probe sets were optimized for annealing temperature and primer-probe concentrations to identify minimum reagent parameters. The selected primer-probe sets were highly sensitive, with DNA concentrations as low as 1 ng adequate for positive species identification. A decoder algorithm was developed based on the cumulative qPCR results that allowed for full automation of species identification. Blinded experiments revealed that the combination of the species-specific primer/probes sets with the automated species decoder resulted in target species identification with 100% accuracy. We also conducted a cost/time comparison analysis between the qPCR assays established in this study with other species identification methods. The qPCR technique was the most cost-effective and least time consuming method of species identification. In summary, probe-based multiplex qPCR assays provide a rapid and accurate method for freshwater fish species identification, and the methodology established in this study can be utilized for various other species identification initiatives.

Keywords: TaqMan probes, Real-Time PCR, species identification, multiplex PCR, freshwater fish species, environmental monitoring

Introduction

Rapid and accurate species identification techniques are important for environmental monitoring programs, specifically in industrial settings. Large scale operations such as power plants, paper mills, petrochemical plants, petroleum refineries, natural gas and food processing plants negatively impact surrounding bodies of water due to use of once-through cooling systems. These system uptake large volumes of lake water where it absorbs heat from the condensers as it circulates throughout the plant before it is released back into the lake as a warmed surface effluent (40). Surrounding wildlife become entrained or impinged at the intake sites of these systems (2). Smaller fishes, such as embryo, larval and juveniles are more susceptible to these industrial processes because they have limited swimming capabilities and therefore cannot avoid the water intake systems (2).

Morphological based species identification technique has often been the traditional method used by environmental monitoring programs to identify impinged fish species at intake points in once-through cooling operations. However, this technique has been proven to have several limitations, specifically with impinged specimens that are often damaged or missing key identification traits. Incorrect identification occurs due to intraspecific phenotypic plasticity of a trait or cryptic variation in both adult and larval fish (3–7). In addition, key morphological traits for identification may only be effective for a particular gender or life stage. This is frequently an issue for larval fish identification as many species share similar morphology (3–6). As well, species can rapidly change during development and often individuals from the same species can appear quite different (3–6). Ko *et al.* 2013 compared the accuracy of morphological identification techniques of larval fish between five different laboratories and found the accuracy rates to be very low; 80.1% accuracy at the family-level, 41.1% accuracy for members among the same genus and 13.5% accuracy at the species-level. Consequently, a high level of expertise is required for morphological identification of larval fish, and often recommended that identification be conservative and limited to the family-level (4). However, ecological and environmental studies necessitate a high level of specificity and accuracy, requiring individuals to be identified to genus- and species-level (5).

Over the last decade, it has been shown that a 650-bp segment of the 5' region of the mitochondrial cytochrome c oxidase I (COI) gene can serve as a universal barcode for the animal kingdom (3,5,7–9). DNA barcoding identifies species through molecular tags based on short regions of DNA from standardised region of the genome (COI, *cytochrome b*, 16S ribosomal RNA, etc) (3,8). DNA barcoding has proven to be a beneficial identification technique for the ecology of cryptic communities, forensic science, monitoring of invasive species and food traceability (3,7,8,11,12). However, DNA barcoding requires several post-PCR processing steps (ethidium bromide staining, gel electrophoresis, PCR purification, etc), which increases processing cost, time and the chance of product contamination.

Real-Time PCR (qPCR) coupled with Taqman probes provides an accurate and high-throughput alternative for species identification when compared to DNA barcoding (9,16–19). The Taqman probe is dual-labelled with a fluorescent reporter dye (e.g FAM, VIC) and a light-absorbing quencher dye (e.g TAMRA, QSY). The probe is unable to emit fluorescent signal when the reporter and quencher dyes are in close proximity. During PCR amplification the probe is hydrolyzed during the extension phase by Taq DNA polymerase. This results in the separation of the dyes and an increase in reporter fluorescence. Following repeated PCR cycles, amplification of the PCR product is exponential resulting in further increase in intensity of the fluorescent signal (17,18,20). When primers and probes are designed to be species-specific, fluorescence occurs only when the forward primer, reverse primer and probe hybridize to the target species DNA (17,20). Here, the TaqMan probe provides a further layer of specificity which is absent in SBYR Green PCR technology resulting in reduced false-positive amplifications (41). Furthermore, compared to DNA barcoding, post-PCR processing steps such as ethidium bromide staining and gel electrophoresis are no longer necessary because the amplification is detected by fluorescent probes (16–18). This not only increases efficiency but also decreases the chance of product contamination (17). Most importantly, probe-based qPCR provides the opportunity to process multiple samples in a single reaction (multiplex). Here, the TaqMan probes can be linked to a variety of fluorescent dyes that have distinct excitation-emission spectra. Multiplex capable qPCR machines are able to acquire simultaneous emission readings, which allows multiple different samples to be processed within a single

well (9,16,17). Probe-based qPCR assays have successfully been developed for numerous applications including detection of bacterial and pathogen strains (17,18), food traceability (9,19,22) and fish identification (16,20,21).

In this study we were interested in developing species-specific qPCR assays for fish species in the Great Lakes, which provides habitat to over 170 species of fish. These fish are not only ecologically important for their surrounding ecosystems but many are of great economical importance for Canada's commercial and recreational fishing industries. In 2005, it was estimated that commercial fisheries in the Great Lakes produced over 35 million dollars in Ontario alone (42). Lake whitefish (*Coregonus clupeaformis*), yellow perch (*Perca flavenscens*) and rainbow smelt (*Osmerus mordax*) represent three of the top eight most harvested fish species in the Great Lakes with values of 8.4 million, 7.9 million and 1.1 million dollars (CAD) respectively (43). Yellow perch had the greatest value per pound with a worth of over $2 per pound (43). Lake whitefish is also an important sustenance fishery for the surrounding Aboriginal communities. Sport fishing is another multibillion dollar industry that relies heavily on the productivity of the Great Lakes. The Great Lakes recreational fishing has an estimated annual net value range of 393 million to 1.47 billion dollars (44). Furthermore, yellow perch, smallmouth bass (*Micopterus dolomieu*) and brook trout (*Salvelinus fontinalis*) are popular fish species targeted by anglers in the Great Lakes (45). Studies have estimated that anglers pay 40 to 55 dollars per day for these species (44).

The objective of this study was to develop and validate species-specific probe-based qPCR assays for the identification of eight freshwater fish species for industrial environmental monitoring programs in the surrounding the Great Lakes region. The eight species of interest included lake whitefish (*Coregonus clupeaformis*), yellow perch (*Perca flavescens*), rainbow smelt (*Osmerus mordax*), brook trout (*Salvelinus fontinalis*), smallmouth bass (*Micropterus dolomieu*), round whitefish (*Prosopium cylindraceum*), spottail shiner (*Notropis hudsonius*) and deepwater sculpin (*Myoxocephalus thompsonii*). Although the round whitefish and deepwater sculpin are not economically important, these species were included in this study since round whitefish is highly sensitive to nuclear power plant operations (38) and deepwater sculpin is an endangered species (37). Utilizing the DNA barcoding gene, COI, species-specific primers and

TaqMan probes were developed for the eight species of interest. In order to identify novel primer-probe sets with maximum species-specificity, two separate primer-probe design criteria were employed. Highest ranked primer-probe sets from both methods were assayed to identify sequences that demonstrated highest specificity. Specificity was determined using control species from same genus and non-target species from different genus. Temperature, primer and probe concentrations were optimized for selected primer-probe sets before they were comprehensively validated through numerous qPCR reactions. A species decoder algorithm was developed based on the qPCR results to automatically identify species thereby allowing full automation of results analysis. Blinded experiments revealed that the combination of the species-specific primer/probes sets with the automated species decoder resulted in target species identification with 100% accuracy. We also conducted a cost/time comparison analysis between the qPCR assays established in this study with other species identification methods.

Materials and Methods

Sample Collection & DNA Extraction

Samples were provided from several external sources outlined in S1Table. DNA was extracted from individual fish muscle tissue, fin clips or liver samples (S1 Table). Extractions were performed using the Qiagen DNEasy Blood and Tissue Kit (Mississuaga, ON) following manufactures guidelines. DNA was eluted from the columns using MilliQ grade nuclease-free water. DNA concentrations, 260:280 and 260:230 ratios were measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific).

DNA Barcoding

Extracted DNA was diluted to 35ng/µL. Polymerase chain reactions (PCR) were run using the universal fish primers Fish F1 (5'-TCA ACC AAC CAC AAA GAC ATT GCC AC-3') and Fish R1 (5'-TAG ACT TCT GGG TGG CCA AAG AAT CA-3') which amplified a 658 base pair region of the COI gene (46). PCR reactions consisted of a total volume of 25 µL with 12.5 µL of iQ Supermix (BioRad

Laboratories, Mississauga, ON), 0.25 µL of 0.1 µM of each primer and 7 ng of template DNA. PCR reactions were run on MJ Mini Personal Thermal Cycler (BioRad Laboratories, ON) with thermal cycling regime of 2 minutes at 94 °C, 35 cycles of 30 seconds at 94°C, 40 seconds at 52°C, 1 minute at 72°C and a final extension of 72°C for 10 minutes. All PCR products were verified on a 1% agrose gel (BioRad Laboratories, ON).

Successfully amplified DNA PCR products were purified using the Qiagen MinElute PCR Purification Kit (Mississauga, ON) following manufactures guidelines. Final elutions were carried out using MilliQ grade nuclease-free water with final elution volume of 10 µL. 0.7 µL of 5 pmol of forward primer was added to 7 µL of purified product. Samples were sequenced using the Sanger method at The Centre for Applied Genomics at SickKids (Toronto, ON).

Specimens were identified by comparing COI sequences to the BLAST database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Species were considered a match when there was ≥98% similarity to an individual species coupled with a bit score of ≥1000.

qPCR

Selected primers were purchased from Sigma-Aldrich Canada Co. (Oakville, ON). Optimal primer annealing temperature and species specificity were validated using QuantStudio5 Real-Time PCR System (Applied Biosystems by Thermo Fischer, ON). SYBR-green based qPCR reactions were prepared in 15 µL volumes containing 2X SensiFAST Sybr Lo-Rox Mix (Bioline, Boston, MA), 300 nM of forward and reverse primer, diethyl pyrocarbonate (DEPC) water and 100 ng of DNA. qPCR reactions were performed in duplicate using the parameters of (1) 95°C for 2 minutes, (2) 95°C for 10 seconds, followed by (3) 58°C for 10 seconds and (4) 72°C for 20 seconds. Steps 2 – 4 were repeated for 40 cycles followed by a melt curve analysis.

TaqMan probes from selected primer sets were ordered from Applied Biosystems, Thermo Fischer Scientific (Foster City, CA). Fluorescents dyes chosen for the TaqMan probers were FAM (~517nm), VIC (~551nm), ABY (~580nm) and JUN (~617nm). Probes were purchased with QSY

quenchers and HPLC purified in unit size of 6000 pmol (1xTE/100pmol format). Primers and TaqMan probe optimization and validation were performed using QuantStudio5 Real-Time PCR System. Taqman probe-based qPCR reactions were prepared in 20 µL volumes containing 100ng of DNA, 50nM to 500nM of forward and reverse primers, 100nM to 200nM of TaqMan probe, 10 µL TaqMan Multiplex Master Mix (Applied Biosystems, CA) and DEPC water. Reactions were run in MicroAmp Optical 96-well reaction plates (Applied Biosystems, CA) using the parameters of 95°C for 20 seconds, followed by 95°C for 1 seconds, 60°C for 20 seconds. This was repeated for 40 cycles. Optimal primer annealing temperature was obtained by performing a temperature gradient of 54°C to 64°C during the annealing stage. Primers and probe concentrations were also extensively optimized. Forward and reverse primers were prepared in varying concentrations of 50, 100, 200 and 400nM and TaqMan probes in concentrations of 100, 200 and 250nM. Optimal primer and probe concentrations were determined based on amplification with the lowest Ct value.

TaqMan probe sets were validated through single and multiplexing reactions with DNA from species of interest and their corresponding control species (same genus) or non-target species (different genus). Single-plexing reactions had 100 ng of DNA from a single species of interest and the corresponding primer-probe set. Multiplexing reaction had four primer-probe sets multiplexed together in a single well with either single or multiple species of DNA. Species detection limits and specificity were validated through 10-fold 8-point standard curves with DNA concentrations of 0.001ng to 300ng. All validation reactions were run in duplicate.

Sequence Alignment

COI sequences were collected for the target species and control species from the same genus (CON) from Barcode of Life Database (BOLD) (http://www.boldsystems.org/). The sequences were aligned using a multiple sequence alignment software (T-Coffee) (https://www.ebi.ac.uk/Tools/msa/tcoffee/). The selected primer-probe sets were analyzed to ensure that there were adequate mismatches between the primer-probe sequences and the corresponding regions on

the COI gene of the control species from the same genus (CONs). Mismatches were calculated as number of base pair (bp) differences between the CON sequences and the primer-probe sequences. Bp differences were highlighted and summed for each primer-probe set.

Results

Species of Interest and Establishment of Control Species Using Geographical and Homology Analysis

The objective of this study was to develop species-specific probe-based qPCR assays for eight freshwater fish species found in the Great Lakes: lake whitefish (*Coregonus clupeaformis*), yellow perch (*Perca flavescens*), rainbow smelt (*Osmerus mordax*), brook trout (*Salvelinus fontinalis*), smallmouth bass (*Micropterus dolomieu*), round whitefish (*Prosopium cylindraceum*), spottail shiner (*Notropis hudsonius*) and deepwater sculpin (*Myoxocephalus thompsonii*). In order to determine primer-probe specificity, we established a list of control species that were highly homologous to the species of interest (Table 1). This was achieved by obtaining CO1 gene sequence information for the target species and their respective CONs using BOLD and FishBase (http://www.fishbase.org/). CO1 gene sequences from all CONs were aligned to their respective target to determine bp similarities. Table 1 reports the alignment scores as percent homology of CON species versus target species and lists the CON species based on homology rank. As well, the location of each CON habitat with regards to the Great Lakes was reported in Table 1. This was performed to discredit certain highly similar CON species when not geographical present in the Great Lakes (47).

Table 1. CO1 gene sequence similarity analysis of species of interest and control species from the same genus (CON). Homology score is represented as the number of base pair matches in the CO1 gene of control species compared to species of interest. Bolded names indicate control species used as negative controls in validating the primer-probe sets.

Species of Interest	Control Species (same genus as species of interest)	Present in Great Lakes (Y/N)	CO1 Gene Similarity (Control Species/Species of Interest)	
			Base Pair Match	Percent (%)
Lake Whitefish (*Coregonus clupeaformis*)	**Cisco (*C. artedi*)**	Y	638/652	98%
	Bloater (*C. hoyi*)	Y	638/352	98%
	Kiyi (*C. kiyi*)	Y	636/652	98%
	Blackfin Cisco (*C. nigripinnis*)	Y	634/652	97%
	Shortjaw Cisco (*C. zenithicus*)	Y	636/652	98%
	Arctic Cisco (*C. autumnalis*)	N	639/652	98%
	Atlantic Whitefish (*C. huntsmani*)	N	632/652	97%
	Bering Cisco (*C. laurettae*)	N	637/652	98%
	Broad Whitefish (*C. nasus*)	N	645/652	99%
	Humpback Whitefish (*C. pidschian*)	N	645/648	99%
	Sardine Cisco (*C. sardinella*)	N	640/652	98%
Smallmouth Bass (*Micropterus dolomieu)*	**Largemouth Bass (*M. salmoides*)**	Y	600/652	92%
Deepwater Sculpin (*Myoxocephalus thompsonii*)	Fourhorn Sculpin (*M. quadricornis*)	N	649/652	99%
Spottail Shiner (*Notropis hudsonius*)	**Pugnose Shiner (*N. anogenus*)**	Y	586/648	90%
	Emerald Shiner (*N. atherinoides*)	Y	590/651	91%
	Bridle Shiner (*N. bifrenatus*)	Y	579/651	89%
	Ghost Shiner (*N. buchanani*)	Y	594/651	91%
	Blackchin Shiner (*N. heterodon*)	Y	585/648	90%
	Blacknose Shiner (*N. heterolepis*)	Y	593/651	91%
	Silver Shiner (*N. photogenis*)	Y	590/651	91%
	Rosyface Shiner (*N. rubellus*)	Y	585/651	90%
	Sand Shiner (*N. stramineus*)	Y	571/645	89%
	Weed Shiner (*N. texanus*)	Y	570/651	88%
	Mimic Shiner (*N. volucellus*)	Y	590/651	91%
	Bigmouth Shiner (*N. dorsalis*)	N	586/651	90%
	River Shiner (*N. blennius*)	N	579/642	90%
	Carmine Shiner (*N. percobromus*)	N	592/651	91%
Rainbow Smelt (*Osmerus mordax*)	Pacific Rainbow Smelt (*O.dentex*)	N	597/648	92%
Yellow Perch (*Perca flavescens)*	Logperch (*Percina caprodes*)	Y	643/652	83%
	Channel Darter (*Percina copelandi*)	**Y**	**557/652**	**85%**
	Blackside Darter (*Percina maculata*)	**Y**	**553/653**	**85%**
	River Darter (*Percina shumardi*)	Y	551/654	84%
Round Whitefish (*Prosopium Cylindraceum*)	**Pygmy Whitefish (*P. coulterii*)**	Y	583/651	90%
	Mountain Whitefish (*P. williamsoni*)	N	624/651	96%
Brook Trout (*Salvelinus fontinalis*)	**Bull Trout (*S. confluentus*)**	Y	611/652	94%
	Lake Trout (*S. namaycush*)	Y	613/652	94%
	Arctic Char (*S. alpinus*)	N	616/652	94%
	Dolly Varden (*S. malma*)	N	617/652	95%

Table 1 demonstrates that many of the most closely related CONs were not located within the Great Lakes, specifically fourhorn sculpin, broad and humpback whitefish which had 99% similarity match to their respective target species. *Notropis* and *Percina* were the exception with their most closely related CONs located in the Great Lakes. Nonetheless, percent similarity for these CONs were only 91% and 85% respectively, which allows for adequate gene sequence dissimilarity for the design of target-specific primer-probe sets. *Coregonus* was the only genus to exhibit CONs within the Great Lakes that had percent similarity greater than 96%. All other target species had CONs within the Great Lakes with sequence similarity below 94%. Therefore, species of interest had either geographical separation from its most closely related CON or the most closely related CON had a lower than 94% homology, with the Coregonus genus being the exception. Selected CON species utilized as negative controls for the validation of the primer-probe sets are bolded in Table 1.

To ensure samples used in this study were correctly identified a subset of the target and non-target samples were barcoded with the CO1 gene and sequenced using the Sanger method. All DNA barcoded samples had a ≥98% match to their corresponding morphologically identified species.

Design of Primer and Probe

The primer-probe sets for the eight species of interest were designed by targeting the COI gene. This region is well suited for species identification purposes because it is relatively conserved within the same species but highly diverged between species (3,8,9). Furthermore, there are a large number of COI sequences publically available for freshwater fish species on databases such as BOLD and GenBank as a result of the Barcode of Life Initiative (3).

Utilizing BOLD, all available sequences for each of the eight target species were obtained regardless of geographical location in order to avoid intraspecies variation: 260 for lake whitefish, 47 for smallmouth bass, 15 for deepwater sculpin, 61 for spottail shiner, 118 for rainbow smelt, 72 for yellow perch, 30 for round whitefish and 46 for brook trout. COI sequences were also collected for closely

related species of each target species (CON). This included all species within the same genus of the target that are found in Canada (Table 1).

Primer-BLAST software offered by National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) was utilized to obtain PCR primer-probe sequences for the species specific COI gene. The following parameters were applied to the software for designing forward and reverse primer pairs: minimum and maximum PCR product size of 70 to 150 respectively, minimum and maximum primer melting temperature of 57 to 63°C with a maximum temperature difference of 3°C between forward and reverse primer sets. As well, corresponding probe parameters included: minimum and maximum size of 18 and 27 nucleotides respectively and a minimum and maximum primer melting temperature of 57 and 63°C.

Primer pair specificity was determined using two different methods:

Method #1

Primer-BLAST was utilized to obtain the top 10 forward and reverse primer pairs and corresponding probe sequences for the COI gene for each species of interest. Species specificity was determined using COI gene sequences from relevant CONs within the Great Lakes. Each primer set (forward, reverse and probe) identified by Primer-BLAST was compared to each CON COI sequence. Primers that had exact matching sequences were given a score of 1. For example, if both the forward primer and probe had exact complementary binding to a CON gene then a score of 2 was given. This analysis was performed for all relevant CON sequences for all 10 primer sets. Scores from all CON COI sequences were summed and the primer sets that scored zero or lowest were considered species specific. When multiple primer sets scored zero, the primer set with the highest rank from Primer-BLAST was chosen for qPCR analysis.

Method #2

A phylogenetic tree for COI gene sequences for relevant CONs and species of interest were generated using a multiple sequence alignment software (T-Coffee) (48). The cladogram parameter of the T-Coffee software revealed the CON species that were most homologous to the species of interest. The COI sequence for the most closely related CON was input in the primer specificity component of Primer-BLAST software. This feature allowed Primer-BLAST to avoid regions that are shared between the species of interest and the most closely related CON. The top ranking set (forward, reverse and probe) identified by Primer-BLAST by this method was chosen for qPCR analysis.

The forward and reverse primer pairs identified from methods one and two were analyzed using SBYR green based qPCR analysis. The primer pair that demonstrated the lowest Ct value was selected as the ideal candidate for species specific identification (Table 2).

Selection of primer-probe sequences

Primer-probe sequences were designed using two novel approaches as described in the *Materials and Methods* section. Analysis of forward and reverse primer specificity using SYBR green based qPCR analysis with 100 ng of the appropriate DNA sample is shown in S2 Table. Overall, C_q values ranged between 12 – 19 and all reactions demonstrated single melt-curve peaks. Interestingly, C_q values obtained from Methods #1 and #2 were within one C_q value demonstrating that both methods provided primer sequences with similar specificity. Nonetheless, the design methodology that resulted with the lowest C_q value was chosen. Therefore, primer-probe sequences for yellow perch and brook trout were obtained from Method #2, while Method #1 was utilized the remaining six species (Table 2).

In Silico Verification of Primer-Probe Specificity

Due to the high degree of homology between species of interest and their respective CONs (Table 1), the selected primer-probe sets were analyzed to ensure that there were adequate mismatches between the primer-probe sequences and the corresponding regions on the COI gene of the CONs. To

accomplish this, COI sequences for target species and their respective CONs were aligned against the selected primer and probe sequences. Number of bp mismatches from the CON sequences and the primer-probe sequences were summed and highlighted in S2 Table. These alignments revealed that all eight primer-probe sets demonstrated perfect alignment with the corresponding target sequences while numerous mismatches were evident when aligned with the CON sequences or sequences from non-target species. In fact, the majority of the bp differences among non-target species ranged from 10 to 19. In regards to the alignment of the primer-probe sequences with corresponding CON, brook trout, yellow perch, rainbow smelt, spottail shiner, and round whitefish all had bp differences of 6 or greater. Smallmouth bass CON, largemouth bass scored lower with bp differences of 5. Deepwater sculpin and lake whitefish scored the lowest for number of mismatches between their CONs and primer-probe sequences (S2.5 Table). Fourhorn sculpin (deepwater sculpin CON), broad whitefish and sardine cisco (lake whitefish CONs) scored the lowest with bp differences of 1, 1 and 2 respectively. These three species are not located within the Great Lakes (Table 1). More importantly, the majority of the CONs for lake whitefish that are located within the Great Lakes demonstrated differences of 3 to 5bps. It is important to note that though these are small differences in bp, they are spread over the reverse primer and probe sequences, increasing specificity for the primer-probe set (49). Stadhouders *et al.* 2010, demonstrates that a single mismatch has the ability to severely reduce priming efficiency, which is essential for species-specificity. Therefore based on the high number of bp differences between the primer-probe sequences and the corresponding CON regions, this analysis revealed that the selected primer-probe sets target COI regions that are highly species specific.

Probe Specification and Primer-Probe qPCR Optimizations

The eight primer-probes were grouped into two sets of four to maximize the multiplexing capabilities of the qPCR machine while minimizing costs required for reagents. The first set and their corresponding fluorescent dyes included smallmouth bass (FAM), spottail shiner (VIC), round whitefish

(ABY) and brook trout (JUN), and the second set comprised of lake whitefish (FAM), deepwater sculpin (VIC), rainbow smelt (ABY) and yellow perch (JUN).

Temperature optimization experiments under single-plex conditions revealed that the primer-probe sets amplified target regions with equal efficiencies at annealing temperatures between 54-64°C; however 60°C was chosen as the optimal temperature for all primer-probe sets based on the manufacture's recommended temperature for the QSY quenchers. Concentration optimization runs established that the primer-probe sets for deepwater sculpin, smallmouth bass and spottail shiner performed most efficiently at lower concentrations ranging from 50 and 100nM. Whereas round whitefish, brook trout, and rainbow smelt had higher optimal concentrations for their primer sets ranging from 400 and 500nM. Table 2 lists the primer/probe sequences (5' to 3'), the respective fluorescent probe dye, optimal primer/probe concentrations and annealing temperature, and the resulting amplicon size.

Table 2. Selected primer and probe sequences targeting the CO1 gene. The table lists the primer/probe sequences (5' to 3') for qPCR analysis, the respective fluorescent probe dye, optimal primer/probe concentrations and annealing temperature, and the resulting amplicon size.

Target Species	Fluorescent Dye	Primer/ Probe	Sequence	Optimal Concentrations (nM)	Optimal Temperature (˚C)	Amplicon Size
Smallmouth Bass	FAM	Forward	TCTTTCCTTCTCCTGCTCGC	50	60	147
		Reverse	GGAGACACCCGCAAGATGAA	50		
		Probe	GCTGGAGCTGGCACTGGGTG	100		
Spottail Shiner	VIC	Forward	CTATTATTAGCTTCTTCTGGGGTTG	50	60	105
		Reverse	GAGGTCTACTGATGCGCCC	50		
		Probe	GCAGGCAATCTTGCCCACGC	100		
Round Whitefish	ABY	Forward	AATGTAATCGTCACGGCCCA	500	60	125
		Reverse	CGGGGGAATGCTATATCGGG	500		
		Probe	TGACTAATTCCCCTTATGATCGGAGCA	100		
Brook Trout	JUN	Forward	CGGTACGGGGTGAACAGTTT	400	60	103
		Reverse	GGAAATGCCAGCTAAATGTAGGG	400		
		Probe	CTCGCCCACGCAGGAGCTTC	200		
Lake Whitefish	FAM	Forward	TCTCCCTCCACTTAGCTGGT	200	60	118
		Reverse	GCCCAGACAAAAAGAGGGGT	400		
		Probe	TTCCTCTATCTTGGGGGCCGTT	200		

Deepwater Sculpin	VIC	Forward	CTTAGCCTCTTCGGGGGTTG	100	60	148
		Reverse	TGCTCCGAGGATCGAAGAGA	100		
		Probe	CCACGCGGGAGCCTCTGTTG	100		
Rainbow Smelt	ABY	Forward	CGATTATGATCGGCGGGTTTG	400	60	76
		Reverse	ATGCGAGGGAAGGCCATATC	400		
		Probe	CCCCCTTATGATTGGGGCCCCA	200		
Yellow Perch	JUN	Forward	GATCGGTGCCCCTGACATAG	200	60	146
		Reverse	TCCCAGCAAGAGGGGGATAA	400		
		Probe	AAGCCGGAGCTGGTACCGGA	200		

Determination of Primer-Probe Sensitivity Under Single-plex vs Multiplex qPCR Conditions

In order to establish the DNA detection limits of the primer-probe sets, single-plex qPCR reactions were performed using target species DNA concentrations ranging from 0.001ng to 300ng. All primer-probe sets were able to amplify their target species at very low concentrations (Figure 1 and S3 Table). In detail, primer-probe sets for spottail shiner, brook trout, lake whitefish, yellow perch and deepwater sculpin were able to detect the respective target species at DNA concentrations as low as 0.001ng. Rainbow smelt primers-probe demonstrated amplification at DNA concentrations of 0.003ng, while smallmouth bass and spottail shiners were effective as low as 0.1ng DNA. Taken together, the primer-probe sets demonstrated excellent detection limits revealing that the sets are highly sensitive.

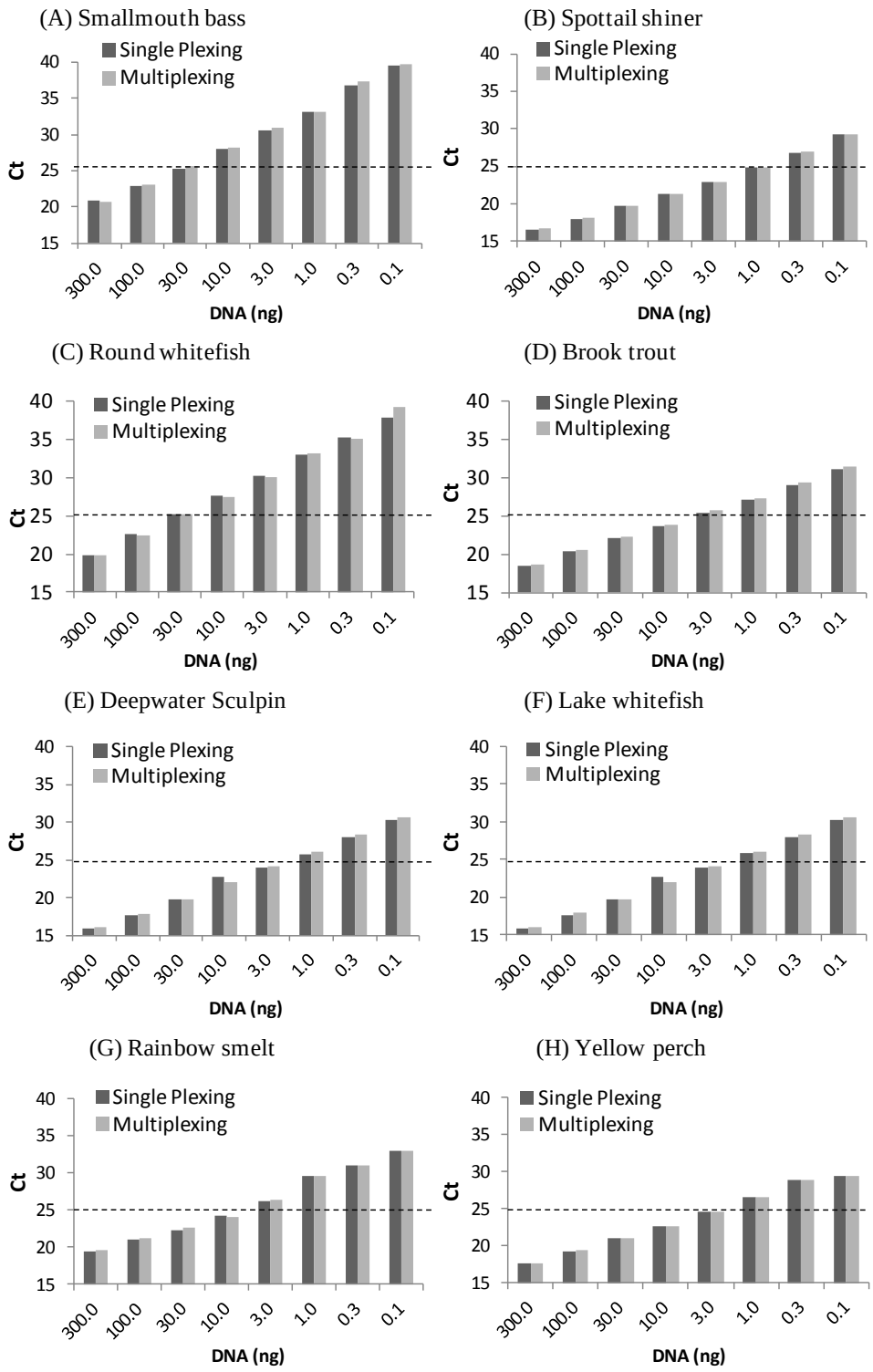

Figure 1. DNA concentration versus qPCR Ct values for each of the species-specific primers/probe sets. Data represents DNA concentrations from 0.03 ng to 300 ng under single-plex and multiplex qPCR conditions. Sensitivity of primer-probe set highlighted by the species-specific Ct cut-off value of 25 (dotted line).

To determine the effects of multiplex analysis on the sensitivity of the primer-probe sets, qPCR runs were performed using target species DNA (0.001ng to 300ng) under multiplex conditions. Figure 1 demonstrates that all primer-probe sets demonstrated similar detection limits under both single-plex and multiplex conditions (S3 Table). The Ct values for multiplexing were within 0.1 to 1.3 when compared to their respective single-plex reactions. Therefore, these results demonstrate that the primer-probe set combinations are multiplex compatible.

Determination of Primer-Probe Specificity

Next the specificity of the primer-probe sets were established by performing qPCR reactions using DNA (0.001ng to 300ng) from appropriate CON species (S3 Table). These results demonstrated that the most closely related CONs for smallmouth bass and lake whitefish were undetectable at all concentrations. CONs for round whitefish, yellow perch, spottail shiner and brook trout were only detectable at higher DNA concentrations of 300 and/or 100ng, with significantly higher Ct values than the targets (35.48 to 39.7 for CONs compared to below 20 for target species) (S3 Table). Taken together, the primer-probes sets are ineffective in amplifying CON sequences demonstrating that the sets are highly specific for the intended target species.

A rigorous specificity analysis was conducted by performing all possible combinations of primer-probe sets with 100 ng DNA from species of interest, CONs or non-target species (different genus). At 100 ng, the primer-probes robustly detected their respective target species with Ct values between 16.21±0.35 and 23.79±0.69 (Table 3). More importantly, the majority of CON species and non-target species were undetectable, while others were detected with Ct values greater than 32. These results establish that at 100 ng DNA from a tissue sample, target species can be detected with Ct value below 25. Therefore by setting a Ct cut-off value of 25 or below, these results suggest that target species can be identified with 100% accuracy and complete absence of false-positive detection.

Table 3. qPCR Ct values for species of interest, control species (same genus) and non-target species (different genus) for each species-specific primers/probe set. 100ng of DNA was used in all reactions. Species that were undetectable were marked as "-".

Probe Set 1					Probe Set 2			
Primer/ Probe Set	**Species**	**# of Samples**	**Average Ct Value**		**Primer/ Probe Set**	**Species**	**# of Samples**	**Average Ct Value**
Smallmouth Bass	Smallmouth Bass	10	23.79 ± 0.69		**Lake Whitefish**	Lake Whitefish	5	16.21 ± 0.35
	Spottail Shiner	10	-			Deepwater Sculpin	5	-
	Round Whitefish	10	-			Rainbow Smelt	14	-
	Brook Trout	4	-			Yellow Perch	10	-
	Lake Whitefish	5	-			Smallmouth Bass	10	37.66 ± 0.21
	Deepwater Sculpin	5	-			Round Whitefish	10	37.71 ± 0.11
	Rainbow Smelt	14	-			Spottail Shiner	10	-
	Yellow Perch	10	-			Brook Trout	4	-
	Largemouth Bass	6	-			Cisco	2	-
Spottail Shiner	Spottail Shiner	10	22.61 ± 1.14			Bloater	1	-
	Smallmouth Bass	10	-			Kiyi	1	-
	Round Whitefish	10	-		**Deepwater Sculpin**	Deepwater Sculpin	5	18.13 ± 0.43
	Brook Trout	4	-			Lake Whitefish	5	-
	Lake Whitefish	5	-			Rainbow Smelt	14	-
	Deepwater Sculpin	5	-			Yellow Perch	10	-
	Rainbow Smelt	14	-			Smallmouth Bass	10	-
	Yellow Perch	10	-			Round Whitefish	10	-
	Bigmouth Shiner	2	-			Spottail Shiner	10	-
	Carmine Shiner	3	-			Brook Trout	4	-
	Rosyface Shiner	2	-		**Rainbow Smelt**	Rainbow Smelt	14	18.83 ± 0.38
	Weed Shiner	3	-			Lake Whitefish	5	-
Round Whitefish	Round Whitefish	10	19.79 ± 1.18			Deepwater Sculpin	5	-
	Smallmouth Bass	10	-			Yellow Perch	10	-
	Spottail Shiner	10	-			Smallmouth Bass	10	-
	Brook Trout	4	-			Round Whitefish	10	35.83 ± 0.49
	Lake Whitefish	5	-			Spottail Shiner	10	-
	Deepwater Sculpin	5	-			Brook Trout	4	32.66±0.039
	Rainbow Smelt	14	-		**Yellow Perch**	Yellow Perch	10	19.12 ± 0.37
	Yellow Perch	10	36.35 ± 0.46			Lake Whitefish		-
	Pygmy Whitefish	3	35.5 ± 0.43			Deepwater Sculpin		
Brook Trout	Brook Trout	4	16.74 ± 0.65			Rainbow Smelt		-
	Smallmouth Bass	10	-			Smallmouth Bass		-
	Spottail Shiner	10	-			Round Whitefish		36.36 ± 0.46
	Round Whitefish	10	-			Spottail Shiner		35.36 ± 0.91
	Lake Whitefish	5	32.61 ± 0.15			Brook Trout		-
	Deepwater Sculpin	5	-			Log Perch	3	-
	Rainbow Smelt	14	-			Blackside Darter	2	39.04 ± 1.96
	Yellow Perch	10	-			River Darter	2	-

	Lake Trout	2	39.64 ± 1.36					
	Arctic Char	2	38.49 ± 0.19					
	Bull Trout	2	36.74 ± 0.52					
	Dolly Varden	2	39.36 ± 0.25					

Blinded Analysis Utilizing Automated Data Decoder Algorithm

A species decoder algorithm was developed to automatically analyze the qPCR data for full automation of species identification. In order to standardize DNA input, 100 ng of sample DNA was chosen for all reactions. Based on the cumulative qPCR results, the following parameters obtained from the QuantStudio software (ThermoFisher) were incorporated into the algorithm. The first parameter included a species-specific Ct cut-off value of 25, as demonstrated in Table 3; detection of all target species occurs below this value and any false detection of non-target species is well above this value (Table 3). Next we utilized the "amplification status" parameter obtained from the QuantStudio software. In certain qPCR reactions, phantom signals are inappropriately associated with a Ct value. The QuantStudio software utilizes proprietary algorithms to measure amplification of the target sample and an endogenous control target (e.g., β- actin, GAPDH, 18S ribosomal RNA, etc.) in a reference sample and target samples to determine whether this amplification is true or false (50). Therefore, only true amplification reactions as determined by QuantStudio software were included in the analysis. Finally, the "normalized reporter value" (ΔRn) of greater than 0.3 was the final parameter included in the algorithm. ΔRn is also used to avoid phantom signals and is calculated by the magnitude of fluorescent signal over the background noise of the dye. Therefore, if a signal had positive amplification status and ΔRn value of 0.3 or greater the algorithm deemed the signal as a positive amplification.

To ensure reliability and accuracy of the automated species decoder algorithm and the primer-probe sets, a comprehensive blinded study was performed that included all species of interest and their corresponding CONs. Table 4 demonstrates that all species of interest were successfully identified by their corresponding primer-probe sets when performed in a blinded fashion (Table 4). Furthermore, the species decoder algorithm had zero percent false positive detection of CON and non-target species (Table

4). As a whole, the blinded experiments revealed that the combination of the species-specific primer/probes sets with the automated species decoder resulted in target species identification with 100% accuracy.

Table 4. Blinded experiments revealed that the species-specific primer/probe sets identified target species with 100% accuracy. A comprehensive blinded study was performed using the species-specific primer-probe sets in combination with randomized samples consisting of species of interest, control species (same genus) and non-target species (different genus). An automated species decoder algorithm was employed for species identification based on qPCR values. The algorithm utilized species-specific Ct value cut-off of 25 and a positive amplification signal with a normalized reporter value (ΔRn) greater than 0.3. All qPCR reactions used 100 ng of sample DNA.

Probe Set 1					Probe Set 2			
Primer Set	**Species**	**# of samples**	**# of positive IDs**		**Primer Set**	**Species**	**# of samples**	**# of positive IDs**
Smallmouth Bass	Smallmouth Bass	10	10		**Lake Whitefish**	Lake Whitefish	5	5
	Largemouth Bass	6	0			Cisco	2	0
	Spottail Shiner	10	0			Bloater	1	0
	Spottail Shiner CONs	14	0			Kiyi	1	0
	Round Whitefish	11	0			Blackfin Cisco	1	0
	Pygmy Whitefish	3	0			Deepwater Sculpin	5	0
	Brook Trout	4	0			Fourhorn Sculpin	1	0
	Brook Trout CONs	8	0			Rainbow Smelt	14	0
Spottail Shiner	Spottail Shiner	10	10			Yellow Perch	10	0
	Weed Shiner	3	0			Yellow Perch CONs	7	0
	Carmine Shiner	3	0		**Deepwater Sculpin**	Deepwater Sculpin	5	5
	Rosyface Shiner	2	0			Fourhorn Sculpin	1	0
	Bigmouth Shiner	2	0			Lake Whitefish	5	0
	Mimic Shiner	1	0			Lake Whitefish CONs	5	0
	Silver Shiner	1	0			Rainbow Smelt	14	0
	Sand Shiner	1	0			Yellow Perch	10	0
	Blackchin Shiner	1	0			Yellow Perch CONs	7	0
	Pugnose Shiner	1	0		**Rainbow Smelt**	Rainbow Smelt	14	14
	River Shiner	1	0			Deepwater Sculpin	5	0
	Smallmouth Bass	10	0			Fourhorn Sculpin	0	0
	Largemouth Bass	6	0			Lake Whitefish	5	0
	Round Whitefish	11	0			Lake Whitefish CONs	5	0
	Pygmy Whitefish	3	0			Yellow Perch	10	0
	Brook Trout	4	0			Yellow Perch CONs	7	0
	Brook Trout CONs	8	0		**Yellow Perch**	Yellow Perch	10	10
Round Whitefish	Round Whitefish	11	11			Log Perch	3	0
	Pygmy Whitefish	3	0			Blackside Darter	2	0
	Smallmouth Bass	10	0			River Darter	2	0
	Largemouth Bass	6	0			Rainbow Smelt	14	0
	Spottail Shiner	10	0			Deepwater Sculpin	5	0
	Spottail Shiner CONs	14	0			Fourhorn Sculpin	0	0
	Brook Trout	4	0			Lake Whitefish	5	0
	Brook Trout CONs	8	0			Lake Whitefish CONs	5	0

Brook Trout	Brook Trout	4	4					
	Arctic Char	2	0					
	Bull Trout	2	0					
	Lake Trout	2	0					
	Dolly Varden	2	0					
	Smallmouth Bass	10	0					
	Largemouth Bass	6	0					
	Round Whitefish	11	0					
	Pygmy Whitefish	3	0					
	Spottail Shiner	10	0					
	Spottail Shiner CONs	14	0					

Multiplexing of Multiple Species DNA

For further validation and economical purposes we were interested in determining whether the primer-probe sets were able to identify their corresponding species of interest when multiple different species DNA was present in a single reaction. When multiplexed with DNA from multiple different species in a single well, all species of interest were correctly identified by their corresponding primer-probe sets (Table 5). DNA detection limit analysis showed that spottail shiner and deepwater sculpin were detected with as low as 0.001ng of DNA. The majority of primer sets detected their corresponding target species to concentrations of 0.03ng excluding smallmouth bass and round whitefish which detected to 1.0ng and 0.1ng respectively. At a 100ng DNA input, spottail shiner, brook trout, lake whitefish, deepwater sculpin, rainbow smelt and yellow perch primer-probe sets were able to detect their corresponding target DNA within the species decoder Ct cut-off of 25 (Table 5). Round whitefish and smallmouth bass however had Ct values slightly above this cut-off with values of 25.23±0.22 and 28.71±0.058 respectively. Therefore, if multiple species samples were mixed and unable to be separated, species can be identified using the species decoder; however Ct cut-off values would need to be adjusted to account for smallmouth bass and round whitefish primer-probe sets. Taken at large, the overall results demonstrate that the probe-based multiplex qPCR assays developed in this study are highly sensitive and robustly accurate.

Table 5. Primer/probe sets specifically amplify target species despite presence of DNA from multiple species. Table demonstrates Ct values for qPCR experiments performed using 100 ng of DNA for each species listed. Data represented as mean Ct value ± standard error of mean.

Probe Set 1				Probe Set 2		
Probe Set	**Species Sample**	**Ct**		**Probe Set**	**Species Sample**	**Ct**
Smallmouth Bass	Smallmouth bass	28.71 ± 0.058		**Lake Whitefish**	Lake Whitefish	21.48 ± 0.065
	Spottail shiner	undetectable			Deepwater Sculpin	undetectable
	Round whitefish	undetectable			Rainbow Smelt	undetectable
	Brook Trout	undetectable			Yellow Perch	undetectable
Spottail Shiner	Spottail shiner	21.37 ± 0.15		**Deepwater Sculpin**	Deepwater Sculpin	21.37 ± 0.027
	Smallmouth bass	undetectable			Lake Whitefish	undetectable
	Round whitefish	undetectable			Rainbow Smelt	undetectable
	Brook Trout	undetectable			Yellow Perch	undetectable
Round Whitefish	Round whitefish	25.23 ± 0.22		**Rainbow Smelt**	Rainbow Smelt	21.48 ± 0.048
	Spottail shiner	undetectable			Deepwater Sculpin	Undetectable
	Smallmouth bass	undetectable			Lake Whitefish	Undetectable
	Brook Trout	undetectable			Yellow Perch	Undetectable
Brook Trout	Brook Trout	23.44 ± 0.11		**Yellow Perch**	Yellow Perch	23.44 ± 0.039
	Spottail shiner	undetectable			Deepwater Sculpin	Undetectable
	Round whitefish	undetectable			Rainbow Smelt	Undetectable
	Smallmouth bass	undetectable			Lake Whitefish	Undetectable

Discussion

Species-specific qPCR primer-probe sets were successfully developed and validated for the eight species of interest: lake whitefish (*Coregonus clupeaformis*), yellow perch (*Perca flavescens*), rainbow smelt (*Osmerus mordax*), brook trout (*Salvelinus fontinalis*), smallmouth bass (*Micropterus dolomieu*), round whitefish (*Prosopium cylindraceum*), spottail shiner (*Notropis hudsonius*) and deepwater sculpin (*Myoxocephalus thompsonii*). The design of novel forward/reverse primers in conjunction with probe sequences permitted for excellent species specificity. Furthermore, the selected primer-probe sets were

multiplex compatible allowing for the development of rapid and high-throughput assays with uncompromised accuracy. Indeed, the development of a fully automated species-decoder algorithm allowed for target species identification with 100% accuracy while completely removing any false-positive detection of non-target species. Most importantly, the species-specific assays were highly sensitive with detection limits as low as 1 picogram of sample DNA.

Fish that are entrained or impinged at once-through cooling system intake zones are often more difficult to identify because of the premature developmental stage (majority are in embryo or larval stage(2)) or because the samples are degraded, damaged and/or missing key identification traits. The developed qPCR primer-probe assays provide several advantages over traditional morphological species identification methods. Here, probe-based qPCR assays are effective at target species identification regardless of gender, life stage, cryptic variation or intraspecific phenotypic plasticity. Furthermore, Ko *et al.* 2013 illustrated that morphological based species identification methods are highly inaccurate and error-prone (4). On the contrary, the incorporation of the automated species decoder algorithm in our probe-based qPCR analysis eliminates false identification errors due to human biases. Another important advantage of probe-based qPCRs is that this assay is more effective than Sanger sequencing based barcoding method when sample DNA is damaged or degraded. Sanger sequencing requires high-grade DNA to achieve successful amplification of the 650bp barcoding region of the COI gene in order to accurately identify individuals. The qPCR assays however only require 70-150bp region of the COI sequence thus allowing degraded samples with greater chances of identification.

In addition to accuracy, large-scale industrial environmental monitoring programs require species identification assays that are cost-effective and high-throughput capable. The probe-based qPCR technique utilized in this study is substantially more cost-effective (Table 6) and time efficient (Table 7) than DNA barcoding and morphological identification methods. To identify an individual sample, qPCR was the cheapest option at $5.82 compared to an external contractor charge of $13 per sample for morphological identification and $18.79 for DNA barcoding, a difference of $7.18 and $12.97 respectively (Table 6). The cost of the primer-probe sets becomes substantially lower as the sample

number increased. For example, when all wells of a 96-format qPCR machine are utilized, 96 samples can be processed for approximately $560, less than half the cost of the other methods (DNA barcoding and morphological identification methods exceed $1200). Cost for qPCR and DNA barcoding were calculated by the reagents, salary for technicians and equipment used to process samples, and morphological identification was the cost charged by the external contractor to Bruce Power for larval fish identification. Similarly, analysis of time requirements (Table 7) reveals that qPCR is the most rapid identification technique which is capable of processing hundreds of samples in only a matter of hours. Here, preparation of the samples required approximately 1.5 h, which was considerably less than the other two techniques (Table 7). In addition, DNA barcoding required several post-PCR processing steps (gel electrophoresis and imaging) that further increased the processing time compared to qPCR. Even with the removal of the PCR verification step with gel electrophoresis for a more high-throughput system, qPCR would still be a quicker approach to yielding results. Review of the morphological identification method demonstrated that this technique is highly variable, depending on the ichthyologist's level of expertise, as well as other factors including species, life stage and physical composition of the sample. For example, a larval Coregonine species would require significantly more time and expertise to identify than an adult burbot (*Lota lota*). Furthermore, once-through cooling systems mostly impinge embryo or larval fish where morphological identification is less effective and requires highly specialized taxonomists (2). Taken together, Table 6 and 7 reveals that the probe-based qPCR assay developed in this study is the most cost-effective and high throughput species identification method.

Table 6. Approximate cost analysis for each species identification technique performed in this experiment. Figures are represented in CAD dollars.

	Per 1 Sample	**Per 96 Samples**
Real-Time PCR Assays		
DNA Extraction Kit	$3.39	$325.44
Primers	$0.0003	$0.03
Probes	$0.02	$1.92
TaqMan Multiplex Mix	$1.40	$134.4
Microamp 96-well Plates	$0.01	$1.30
Labour	$1	$100
Total	**$5.82**	**$563.09**
DNA Barcoding		
DNA Extraction Kit	$3.39	$325.44
Primers	$0.0009	$0.08
iQ MasterMix	$2.81	$269.76
Agrose gel and Loading Dye	$3.04	$40.07
PCR Purfication Kit	$2.61	$250.75
Strip Tubes	$0.47	$5.64
Sanger Sequencing	$3.50	$336
Labour	$1	$100
Total	**$18.79**	**$1,319**
Morphological Identification		
Labour	$13 ($10 US)	$1,248 ($960 US)
Total	**$13 ($10 US)**	**$1,248 ($960 US)**

Table 7. Approximate time requirement for each species identification technique. Morphological identification time varies depending on species complexity, species such as *Coregonus* required significantly more time and experience to identify than a burbot.

	Per 96 Samples
Real-Time PCR Assays	
DNA Extraction	Prep: 1 hour
	Incubation: 3-4 hours
Real-Time PCR	Prep: 1 hour
	Run Time: 30 minutes
Analysis	30 minutes
Total	**6 - 7 hours**
DNA Barcoding	
DNA Extraction	Prep: 1 hour
	Incubation: 3-4 hours
PCR	Prep: 30 minutes
	Run Time: 2 hours
Gel Electrophoresis	Prep: 30 minutes
	Run Time: 1-2 hours
PCR Purification	1 hour
Analysis	1 hour
Total	**10 - 12 hours**
Morphological Identification	
Labour	8 - 16 hours
Total	**8 - 16 hours**

Many fish species within the Great Lakes are post-glacial and recently diverged. The *Coregonus* specifically has had a recent evolutionary divergence and exhibits extensive interspecies COI haplotype sharing (7,13,51,52). This genus is extremely difficult to identify morphologically as individuals from the same species exhibit multiple different morphologies within the same lake (52,53). As well, DNA barcoding has had several limitations with identifying individuals from Coregonine to the genus and/or species level due to decreased mitochondrial DNA variation between species (7,13). Furthermore, the

Coregonus genus within Canada demonstrated the greatest percent homology to one another out of all the species of interest, with a 97-99% similarity (Table 1). As a result, Lake Whitefish primer-probe set had the lowest number of bp differences between its closely related CON species (Supplementary Table 2.5). Overdyk *et al.* 2016 previously designed a qPCR assay to distinguish Lake Whitefish from the other *Coregonus* species; however, their assay only had one bp difference on their probe sequence for all the *Coregonus* species and had reported non-target Ct detection of 29 for yellow perch. Our assay had 3 to 5bp differences that were spread over the reverse primer and probe sequences and no Ct detection for non-target species. This demonstrates that our assay design criteria was very effective and was able to identify primer-probe sets for even difficult to discriminate species such as the lake whitefish.

The development and validation of the probe-based qPCR assay outlined in this study is translatable to all species. For example, other ecologically and economically important fish species that may be of concern for industrial environmental programs include walleye, lake trout, cisco and round goby (2). As well, the high sensitivity achieved by the qPCR assays demonstrate that the primer-probe sets can be successfully applied to samples of low abundance, including environmental DNA (eDNA) or digested stomach samples. eDNA is novel non-invasive technique used to identify organisms by the fragments of DNA that are released in the environment. Currently, highly expensive next-generation sequencing platform has been used to detect species from low abundant eDNA samples. Our results suggest that the qPCR primer-probe sets developed in this study can potentially offer a more time efficient and cost-effective means to identify specific species from eDNA samples.

Conclusion

Highly accurate and cost-effective species-specific qPCR TaqMan primer-probe sets were successfully developed for the rapid and high-throughput identification of eight ecologically and economically important freshwater fish species. The combination of the species-specific primer/probes sets with an automated species decoder algorithm resulted in target species identification with 100% accuracy coupled with complete absence of false-positive detection from non-target controls. Most

importantly, the probe-based qPCR assays were highly sensitive with detection limits as low as 1 picogram of sample DNA. Furthermore, the probe-based qPCR technique utilized in this study is substantially more cost-effective and time efficient than DNA barcoding and morphological identification methods. In summary, probe-based multiplex qPCR assays provide a rapid and accurate method for freshwater fish species identification, and the methodology established in this study can be utilized for various other species identification initiatives.

Acknowledgments

The researchers would like to thank the following individuals and institutions for providing tissue samples: Dr. Louis Bernatchez and Universite Laval, Dr. Robert Humston and Washington and Lee University, Dr. Nancy Denslow and the University of Florida, Dr. James Johnsons and the Great Lakes Science Centre, Dr. Joanna Wilson and McMaster University, Dr. Hernan Lopez-Fernandez and the Royal Ontario Museum, Dr. Eric B. Taylor and the University of British Columbia, and Dr. Doug Watkinson and the Department of Fisheries and Oceans.

Chapter 3: Identification of aquatic freshwater species using environmental DNA (eDNA)

Identification of aquatic freshwater species using environmental DNA (eDNA)

Emily N. Hulley, Sujeenthar Tharmalingam, Chris Thome, Charles Mitz, and Douglas R. Boreham

(Original Research)

Abstract

Environmental DNA (eDNA) has recently emerged as a highly sensitive, cost effective and non-invasive technique for species detection. Here we utilize previously developed and validated species-specific real-time PCR (qPCR) primer probes to detect yellow perch (*Perca flavescens*) from eDNA samples. In March of 2018 three 1L water samples were collected from six sites across Lake Laurentian (Sudbury, ON). qPCR assays accurately detected yellow perch from the five active sites and not at the control site. qPCR results were validated through sanger sequencing (SS) and next generation sequencing (NGS). Furthermore, the NGS results demonstrated comparable yellow perch detection to those found from the qPCR reactions. The SS and NGS results validate that the yellow perch primer-probe set were highly sensitive and robustly accurate at detecting yellow perch from eDNA samples. In summary, eDNA coupled with the developed qPCR assays can be an accurate and reliable method for the detection of freshwater fish species.

Keywords: environmental DNA (eDNA), sanger sequencing, next generation sequencing, species detection, yellow perch (*Perca flavescens*)

Introduction

Environmental DNA (eDNA) has emerged over the last few years as a highly sensitive, cost effective and non-invasive technique for species detection (23,25,27,28). eDNA are short, species-specific fragments of DNA released from organisms into the environmental through defecation, urination, secretion of mucus or gametes, epidermal cells shedding and carcasses (24,25,28–30). In aquatic environments, eDNA can remain approximately 7 to 21 days depending on conditions (24). Exposure to UV radiation, endogenous nucleases, heat and acidity can degrade eDNA (24,29).

Filtration or precipitation of water samples has been the two main techniques used to collect eDNA (Supplementary Table 4). Precipitation of water samples uses smaller volumes of water (15mL) compared to filtration systems where large volumes of water (≥1L) are vacuum filtered onto micrometer pore-sized filters (Supplementary Table 4) (32,54). Verification of target species eDNA has been analyzed through standard PCR and gel electrophoresis, real-time PCR (qPCR), digital drop PCR (ddPCR), and next generation sequencing (NGS) (Supplementary Table 4). eDNA has been used for the detection and monitoring of numerous aquatic endangered (27,36) and invasive species (23,31,33–35), and more recently to estimate species biomass (25,26,28).

In this study, we were interested in using previously designed qPCR assays from Chapter 2 to detect fish species from eDNA samples. Lake Laurentian, Sudbury (ON, Canada) was selected as our study site and is a popular location for recreational fishermen. The lake contains two species of fish, yellow perch (*Perca flavescens*), our target species, and northern pike (*Esox lucius*) (39). Three 1L water samples were collected from six sites across the lake in March 2018. Yellow perch primer-probe sets detected yellow perch from 5 of the 6 sites. Yellow perch primer-probe results were confirmed using Sanger sequencing (SS) and NGS of each sites eDNA samples. NGS demonstrated analogous results of yellow perch detection at the six sites to those obtained by the qPCR assay. The SS and NGS results validate that the yellow perch primer-probe set were highly sensitive and robustly accurate at detecting yellow perch from eDNA samples.

Methods

eDNA Field Sampling & Extractions

Samples were collected from the study area, Lake Laurentian (Sudbury, ON) in March 2018. Three 1L samples were collected from six sites across the lake (Figure 1). 1L Nalgene bottles were autoclaved and decontaminated with 10% bleach prior to sampling. Holes were drilled through the ice 10 inch in diameter with a motorized auger. A six foot plastic water column was used to sample water from below the ice (~95cm deep). The water column and Nalgene bottles used for each site were rinsed with that sites lake water three times prior to sample collection. Water samples were stored on ice in the field (~2 to 3 hours) and during transportation back to the laboratory where they were stored at -20˚C until filtration and DNA extraction.

Figure 2. Site locations at the study area, Lake Laurentian (Sudbury, ON). 3x1L water samples were collected from each of the six sites.

All equipment used for filtration and extraction were autoclaved and treated with 10% bleach and then exposed to ultraviolet light for 30 minutes before use (forceps, vacuum flask etc). Water samples

were removed from the freezer, thawed, and the exterior of the bottle was treated with 10% bleach and 70% ethanol before being transferred into a sterile biological safety cabinet (BSC). Water samples from each site were inverted several times before being vacuum filtered onto 0.2 μm Gamma Irradiated MicroFunnel Filters with Supor Membrane (PALL Life Sciences, NY). Due to the turbidity of the water samples, 1L samples required two filters (2x500mL) for a total of 6 filters per site. Each filter was then placed into 5mL PowerWater DNA Bead Tube. DNA was extracted using the Qiagen DNeasy PowerWater Kit (Mississuaga, ON**)** following manufactures guidelines. DNA was eluted from the columns using diethyl pyrocarbonate (DEPC) water instead of the sterile elution buffer (EB) solution provided. DNA concentrations were measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific).

Control Samples, Collection & DNA Extraction

Samples used for controls were provided from several external sources outlined in Supplementary Table 1. Yellow perch tissue samples were used as a positive control. Smallmouth bass, lake whitefish and rainbow smelt were used as negative controls. DNA was extracted from individual fish muscle tissue and fin clip samples (Supplementary Table 1). Extractions were performed using the Qiagen DNeasy Blood and Tissue Kit (Mississauga, ON) following manufactures guidelines. DNA was eluted from the column using MilliQ grade nuclease-free water. DNA concentrations, 260:280 and 260:230 ratios were measured using a spectrophotometer.

Real-Time PCR

Primers were purchased from Sigma-Aldrich Canada Co. (Oakville, ON). The TaqMan probe was ordered from Applied Biosystems, Thermo Fisher Scientfic (Foster City, CA). The probe was purchased with JUN fluorescent dye, QSY quencher and HPLC purified unit size of 6000 pmol (1xTE/100pmol format). A QuantStudio5 Real-Time PCR (Applied Biosystems by Thermo Fischer, ON) was used for the yellow perch detection assays. eDNA qPCR reactions were prepared in 20 μL volumes containing 15ng

of DNA, 200nM of forward primer, 400nM of reverse primer, 200nM of TaqMan probe, 10µL TaqMan Multiplex Master Mix (Applied Biosystems, CA) and diethyl pyrocarbonate (DEPC) treated water. Reactions were performed in MicroAmp Optical 96-well reaction plates (Applied Biosystems, CA) using the parameters of 95°C for 20 seconds, followed by 95°C for 1 second, 60°C for 20 seconds. This was repeated for 40 cycles. All eDNA samples, negative and positive controls were run in duplicate.

qPCR products were purified using the Qiagen MinElute PCR Purification Kit (Mississauga, ON) following manufactures guidelines. Final elutions were carried out using DEPC treated water with final elution volume of 10 µL. 0.7 µL of 5 pmol forward primer was added to 7 µL of purified product. Samples were sequenced using the Sanger method at The Centre for Applied Genomics at SickKids (Toronto, ON).

Samples were identified by comparing COI sequences to the BLAST database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Species were considered a match when there was ≥98% similarity to an individual species.

Next Generation Sequencing

Replicate samples from each site were combined to produce a total sample volume of 20µL for each site. Samples were sent to Molecular Research DNA (MR DNA, Shallowater, TX, USA) for NGS on an Illumina HiSeq 2500. Sequence data from NGS were processed using MR DNA ribosomal and functional gene analysis pipeline (www.mrdnalab.com , MR DNA, Shallowater, TX). Barcodes, primers, short sequences (<150bp) and sequences with ambiguous base calls were removed. Operational taxonomic units (OTU) were defined clustering at a 1% divergence (99% similarity) followed by removal of chimeras and singleton sequences. Final OTUs were taxonomically identified using BLAST databases RDPII (http://rdp.cme.msu.edu) and NCBI (www.ncbi.nlm.nih.gov). OTUs were processed into count files that contained the number of sequences for each taxonomic level. Individuals from the count files were considered a true match to the taxonomic level when percent homology was 98% or greater.

Geographic distributions of species were also taken into consideration when identifying individuals; i.e., species that do not occur in the study region were not included for consideration.

Results

Sample Sites

The objective of this study was to identify yellow perch from eDNA samples collected from Lake Laurentian using previously developed species-specific qPCR assay. Water samples were collected from each site in March of 2018 (Figure 2). Depth measurements (total depth, open water depth, ice depth) and sediment composition were also recorded at each site (Table 8). All six sites had soft bottom sediment composition and total ice coverage of approximately 90 to 95 cm (Table 8). Site B had the greatest depth, with a total depth of 3.50m and a water depth below the ice of approximately 2.60 to 2.65m (Table 8). Site A, C and F all had similar water depths of 1.0 to 1.25 m. Site D and E were significantly shallower than the other sites, with water depths of 0.1-0.65m (Table 8). Site E was considered a negative control for yellow perch in this study, as water depth below the ice surface was only 10 to 15 cm. Minimal water depth at this site would likely create anoxic conditions and therefore an unsustainable environment for yellow perch.

Table 8. Lake Laurentian eDNA sampling sites depth and sediment composition. Ice coverage on the lake was approximately 90-95cm in depth.

Site	Total Depth (m)	Water Depth (m)	Sediment Composition
A	1.95	1.05-1.0	soft bottom
B	3.50	2.60-2.55	soft bottom
C	2.20	1.30-1.25	soft bottom
D	1.55	0.65-0.60	soft bottom
E	1.05	0.15-0.10	soft bottom
F	creek opening	1.0	soft bottom

Yellow Perch Detection with qPCR Assay

Detection of yellow perch from the Lake Laurentian eDNA samples and the controls were completed using species-specific RT PCR primer-probe set. qPCR reactions were performed in single-plex reactions with concentrations of eDNA and tissue samples at 15ng. Yellow perch tissue samples were used as a positive control and had an average Ct of 21.91±0.31 (Supplementary Table 5). Smallmouth bass, rainbow smelt and lake whitefish were used as negative controls and were all undetectable (Supplementary Table 4). This indicates that the primer-probe set were functioning accurately and demonstrating strong species-specificity.

Ct values were detected from all eDNA samples. However, these values were above 30, which was higher than those of the positive control (Table 9 and Supplementary Table 5). To avoid any potential false detection due to the higher Ct values from the eDNA samples, we utilized the "amplification status" parameter from the QuantStudio software. In certain qPCR reactions, phantom signals can become incorrectly associated with a Ct value. The QuantStudio software utilizes proprietary algorithms and determines whether this amplification is true or false. Therefore, only true amplification reactions as determined by QuantStudio software were considered a true detection. In addition to the amplification status, we deemed Ct values above 35 to be unreliable, as this out of the range of efficiencies for qPCR machines (55).

At 15ng of DNA yellow perch was detected at sites A, B, C, D and F (Table 9). All qPCR reactions had positive amplification and Ct values below 35. Site A exhibited the lowest Ct from all the sites, with an average Ct of 31.94±0.56 (Table 9). Site E was our negative control, due to the shallow depth being unsustainable for yellow perch. From the obtained qPCR results, yellow perch was not detected in Site E (Table 9). Here, Site E exhibited a Ct value greater than 35 and had an inconclusive amplification status. Therefore, these results indicate that the qPCR assay has the ability to detect its target species from an eDNA sample.

Table 9. qPCR Ct values and amplification status for each site eDNA sample detected by the yellow perch primer-probe sets. 15ng of eDNA was used in all reactions. Next Generation Sequencing counts represent the total number of sequences of yellow perch identified from eDNA samples with ≥98% homology match. Sanger sequencing identities demonstrates the number of base pair matches and percent homology to yellow perch COI sequence.

Site	Real-Time PCR		Next Generation Sequencing	Sanger Sequencing
	Ct value	Amplification Status	Counts	Identities
A	31.94±0.56	positive	181	-
B	32.39±0.34	positive	35	108/109 (99%)
C	34.66±0.57	positive	16	106/107 (99%)
D	32.96±0.17	positive	33	109/109 (100%)
E	37.61±0.58	inconclusive	0	-
F	34.92±0.33	positive	14	108/109 (99%)

Validation with Sanger Sequencing and Next Generation Sequencing

In order to verify the obtained qPCR results for each site, samples were sequenced through SS and NGS methods. Following qPCR reactions, samples were sanger sequenced to confirm obtained Ct values for each of the sites. Site B, C, D and F all had a ≥99% match to the BLAST yellow perch COI sequence (Table 9). Site E qPCR product did not trim to the yellow perch COI sequence and therefore did not have a percent match to any yellow perch sequences on BLAST (Table 9). Site A, however did not sequence correctly, this was likely due to degraded sample and consequently the inability to be sequenced (Table 9). These results confirmed that our obtained Ct values were accurate in identifying the eDNA samples that contained yellow perch from those that did not.

Next, we were interested in seeing if NGS could distinguish a difference in the amount of yellow perch DNA from the different sites, in addition to further validating our qPCR and Sanger sequencing results. From the NGS results, counts were obtained for each of the six sites. These counts represent the

total number of sequences of yellow perch identified from eDNA samples with ≥98% homology match. For each of the six sites, NGS results demonstrated comparable yellow perch detection to those found from the qPCR reactions (Table 9). Sites A, B, C, D and F all had counts with a percent homology match of 98% or greater to yellow perch (Table 9). Site A had the greatest number of counts, 181, which corresponded to our qPCR results with it having the lowest Ct value (Table 9). Site B and D had similar Ct values, as well as, similar NGS counts of 35 and 33 respectively (Table 9). Site C and F demonstrated the lowest amount of yellow perch DNA from both the qPCR and NGS results (Table 9).

Furthermore, Site E exhibited zero sequence counts for yellow perch, which was consistent with the RT PCR results, indicating the absence of yellow perch from the site (Table 9). Taken at large, the overall results from the SS and NGS methods demonstrate that the yellow perch primer-probe set were highly sensitive and robustly accurate at detecting yellow perch from eDNA samples.

Discussion

From the results presented, we demonstrated that eDNA coupled with qPCR assays can be an accurate and reliable method for the detection of freshwater fish species. qPCR assays developed from Chapter 2 accurately detected yellow perch from the sampled sites at Lake Laurentian. qPCR results were validated through both SS and NGS. Furthermore, the NGS results demonstrated comparable yellow perch detection to those found from the qPCR reactions, confirming the sensitivity and accuracy of the qPCR assay detection of yellow perch from eDNA samples.

eDNA coupled with qPCR offers a cost-effective, efficient and accurate method for the detection of specific species, including invasive or rare species, compared to traditional field surveying and identification techniques. Often rare and/or endangered species have very low detection probabilities, especially in aquatic environments where organisms are small and hidden below the water surface (27,56). Fish monitoring programs traditionally use electrofishing gear or nets to estimate occurrence of species within an aquatic environment, which have several disadvantages (27,36,56). First, these tools typically have low capture probability per target organism and are often only reliable indicators for

species that are present in moderate to high abundance (27). For rare species, the low capture probabilities of these tools often lead to inaccurate inference of a species presence in the surveyed site. This requires an increase in sampling effort, which is often not feasible due to cost and time (27). Invasive monitoring programs also require the detection of invasive species while they are still in low abundance to ensure controls and eradication efforts will be successful. However similar to rare species, these programs lack the tools required to detect aquatic invasive species when they are in low abundance and can only detect them once the density has reach a certain threshold in which the species has a well-established population and spread (31,33). In addition, these techniques are highly invasive. Fish traps pose a high risk for by-catch for nontarget species, which if not checked at regular intervals can result in high amounts of fish mortality (ex: gill nets). As well, by-catch of nontarget species can also result in the mortality of the target species. Egelyng Sigsgaard *et al.* 2014 reported that the endangered weather loach is often killed in surveillance traps by predatory by-catch fish such as eels. eDNA coupled with species-specific qPCR primer-probe sets have proved to be an accurate, cost-effective, non-invasive and efficient solution to traditional surveillance methods (23,27,31,36,56). eDNA sampling requires only the collection of a few litres of surface water from each sample site, which eliminates the cost, time and invasiveness of using traditional sampling tools. As well, eDNA requires minimal post processing steps compared to other molecular identification techniques. The extraction of eDNA from filters through a kit is substantially quicker than the extraction of DNA from a tissue, as the lysis step occurs in minutes instead of hours. Species-specific qPCR assays also allow for the rapid analysis of eDNA without the post-processing steps required by other molecular identification techniques (ex: gel electrophoresis and PCR purification with Sanger sequencing). Results can be obtained from the qPCR assays in a matter of hours from highly diluted and/or degraded eDNA samples.

NGS is a revolutionary sequencing technique that can sequence millions of small fragments of DNA in parallel (57). Bioinformatic analyses are then used to piece these fragments together by mapping the individual reads to a reference genome. Each of these bases within the genome are sequenced multiple times, which provides a high depth and accurate results (57). Recently, studies have begun to couple NGS

with eDNA to assess biodiversity (58,59). NGS coupled with eDNA has already demonstrated unparallel barcoding potential without the steps of designing, optimizing and validating qPCR primers for specific species. From our study site, a total of 33 species, with percent homology of ≥98%, were observed from the NGS results (Supplementary Table 6). Many of the species were invertebrates and insects, however, several *Notropis* species, amphibian, reptile and mammal species were also observed. Interestingly, Northern Pike was not detected by NGS, even though the species is known to live in Lake Laurentian. Despite the potential power of NGS for the high-throughput identification of species, there are still some limitations. The first limitation being cost. Many of the NGS platforms (Illumina, Roche 454, SOLiD etc) available are significantly more expensive than those of qPCR or SS. Prices vary depending on the desired number of reads, sample size and bioinformatic analysis used but costs typically average from hundreds to thousands of dollars. Another limitation is the time required to analyze the large datasets that the NGS results produce. A computer with high processing capabilities is required to map the millions of fragment reads back to a reference genome. This also places a reliance on having a high-quality reference database for every organism genome to ensure accurate sequencing (60). Even with these limitations, eDNA coupled with NGS has enormous potential to estimate biodiversity and discover new species (60).

eDNA concentration and detection rates do experience seasonal variation with peak detection exhibited in the summer (July) and lower detection in the winter (January) (61). In future studies, we would recommend that the study site be sampled again during the summer season to compare the qPCR and NGS results collected from the present study. As well, future experiments could design a species-specific qPCR assay for northern pike and examine potential reasons for its absence in the NGS results. In conclusion, it was demonstrated that eDNA coupled with qPCR assays can be an accurate and reliable method for the detection of freshwater fish species. Furthermore, the potential for NGS coupled with eDNA can provided unparallel potential for determining species biodiversity and discovering new species.

Chapter 4: Conclusions and Future Direction

Highly accurate and cost-effective species-specific qPCR TaqMan primer-probe sets were successfully developed for the rapid and high-throughput identification of eight ecologically and economically important freshwater fish species. The combination of the species-specific assays with an automated species decoder algorithm resulted in target species identification with 100% accuracy couple with the absence of false-positive detection from nontarget controls. In addition, the probe-based assay utilized in this thesis was substantially more cost-effective and time efficient than DNA barcoding and morphological identification methods. Most importantly, the primer-probe assays were highly sensitive with detection limits as low as 1 pictogram of sample DNA. TaqMan primer-probe sets were able to accurately detect their corresponding target species from eDNA samples, without false detection from negative controls. The development and validation of the qPCR assays in this thesis are translatable to all species. For example, other ecologically and economically important fish species that may be of concern to industrial environmental programs include emerald shiner, walleye, alewife, lake trout, cisco and round goby (2). These primer-probe sets can be applied in environmental monitoring programs to provide a highly accurate and cost-effective method to identify impacted individuals.

In addition, the primer-probe assays developed in Chapter two proved to be highly sensitive for detection of their target species when samples were diluted and degraded in eDNA samples. Yellow perch primer-probe sets detected yellow perch from the sampled sites at Lake Laurentian. qPCR results for each of the six sites were validated using SS and NGS. NGS demonstrated analogous results of yellow perch detection to those found from the qPCR reactions. The SS and NGS results validate that the yellow perch primer-probe set were highly sensitive and robustly accurate at detecting yellow perch from eDNA samples.

The methodologies designed in this thesis can be applied to environmental monitoring programs to detect, identify and monitor indicator, rare, endangered or invasive species through qPCR, or eDNA coupled with qPCR and/or NGS. The qPCR assays can be used specifically for once-through cooling systems to immediately identify larval and embryo fishes that become impinged or entrained in the system. qPCR assays can be developed for other impacted species, to create an inhouse library for the high-through put and cost-effective identification of impinged or entrained individuals. Future experiments should further examine the potential for eDNA to estimate species biomass through the qPCR methodologies proposed in this thesis and NGS. As well, future studies could examine the applicability of qPCR coupled with eDNA with amphibian or reptile species, as these species have not been as widely researched using these techniques.

References

1. Madden N, Lewis A, Davis M. Thermal effluent from the power sector: An analysis of once-through cooling system impacts on surface water temperature. Environ Res Lett. 2013;8(3).
2. Smith C. 2016 Environmental Monitoring Program Report [Internet]. B-REP-07000-00009 . Bruce Pwer; 2017. Available from: http://www.brucepower.com/wp-content/uploads/2017/05/2016-EMP-Report.pdf
3. Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, Burridge M, et al. Identifying Canadian freshwater fishes through DNA barcodes. PLoS One [Internet]. 2008;3(6):e2490. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22423312
4. Ko HL, Wang YT, Chiu TS, Lee MA, Leu MY, Chang KZ, et al. Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. PLoS One [Internet]. 2013;8(1):e53451. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23382845
5. Valentini A, Pompanon F, Taberlet P. DNA barcoding for ecologists. Trends Ecol Evol [Internet]. 2009;24(2):110–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19100655
6. Teletchea F. Molecular identification methods of fish species: reassessment and possible applications. Rev Fish Biol Fish [Internet]. 2009;19(3):265. Available from: https://doi.org/10.1007/s11160-009-9107-4
7. April J, Mayden RL, Hanner RH, Bernatchez L. Genetic calibration of species diversity among North America's freshwater fishes. Proc Natl Acad Sci U S A [Internet]. 2011;108(26):10602–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21670289
8. Costa FO, Carvalho GR. The Barcode of Life Initiative: synopsis and prospective societal impacts of DNA barcoding of Fish. Genomics, Soc Policy [Internet]. 2007;3(2):29. Available from: https://doi.org/10.1186/1746-5354-3-2-29
9. Rasmussen Hellberg RS, Morrissey MT, Hanner RH. A multiplex PCR method for the identification of commercially important salmon and trout species (Oncorhynchus and Salmo) in North America. J Food Sci [Internet]. 2010;75(7):C595-606. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21535525
10. Meyer CP, Paulay G. DNA barcoding: Error rates based on comprehensive sampling. PLoS Biol. 2005;
11. Andrea Galimberti Alessia Losa, Ilaria Bruni, Silvia Federici, Maurizio Casiraghi, Stefano Martellos, Massimo Labra FDM. DNA barcoding as a new tool for food traceability. Food Res Int. 2012;50(1):55–63.
12. Benedict A. Maralit Minerva Fatimae H. Ventolero, Sweedy Kay L. Perez, Demian A. Willette, Mudjekeewis D. Santos RDA. Detection of mislabeled commercial fishery by-products in the Philippines using DNA barcodes and its implications to food traceability and safety. Food Control. 2013;33(1):119–25.
13. L. Bernatchez JJD and SB. Population bottlenecks: influence on mitochondrial DNA diversity and its effect in coregonine stock discrimination. J Fish Biol. 1989;35(Supplement A):233–44.
14. Rubinoff D, Cameron S, Will K. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. Journal of Heredity. 2006.
15. Galtier N, Nabholz B, Glemin S, Hurst GDD. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol Ecol. 2009;
16. Watanabe S, Minegishi Y, Yoshinaga T, Aoyama J, Tsukamoto K. A quick method for species identification of Japanese eel (Anguilla japonica) using real-time PCR: an onboard application for use during sampling surveys. Mar Biotechnol [Internet]. 2004;6(6):566–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15690101

17. Weller SA, Elphinstone JG, Smith NC, Boonham N, Stead DE. Detection of Ralstonia solanacearum strains with a quantitative, multiplex, real-time, fluorogenic PCR (TaqMan) assay. Appl Env Microbiol [Internet]. 2000;66(7):2853–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10877778
18. del Cerro A, Mendoza MC, Guijarro JA. Usefulness of a TaqMan-based polymerase chain reaction assay for the detection of the fish pathogen Flavobacterium psychrophilum. J Appl Microbiol [Internet]. 2002;93(1):149–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12067384
19. Kesmen Z, Gulluce A, Sahin F, Yetim H. Identification of meat species by TaqMan-based real-time PCR assay. Meat Sci [Internet]. 2009;82(4):444–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20416686
20. Keith M. Bayha Frank J. Hernandez Jr WMG. Multiplex assay to identify eggs of three fish species from the northern Gulf of Mexico, using locked nucleic acid Taqman real-time PCR probes. Aquat Biol. 2008;4:65–73.
21. Martini I. Taylor Ines Rico, Ciro Rico CF. Species-specific TaqMan probes for simultaneous identification of (Gadus morhua L.), haddock (Melanogrammus aeglefinus L.) and whiting (Merlangius merlangus L.). Mol Ecol Notes. 2002;2:599–601.
22. Trotta M, Schonhuth S, Pepe T, Cortesi ML, Puyet A, Bautista JM. Multiplex PCR method for use in real-time PCR for identification of fish fillets from grouper (Epinephelus and Mycteroperca species) and common substitute species. J Agric Food Chem [Internet]. 2005;53(6):2039–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15769133
23. Gleeson EMF and D. Environmental DNA detection of redfin perch, Perca fluviatilis. Conserv Genet Resour. 2016;8:115–8.
24. David S. Pilliod Matthew B. Laramine, Lisette P. Waits CSG. Application of environmental DNA for inventory and monitoring of aquatic species [Internet]. United States Geological Survey; 2013. Available from: https://pubs.usgs.gov/fs/2012/3146/pdf/fs2012-3146.pdf
25. Maruyama A, Nakamura K, Yamanaka H, Kondoh M, Minamoto T. The release rate of environmental DNA from juvenile and adult fish. PLoS One [Internet]. 2014;9(12):e114639. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25479160
26. Doi H, Uchii K, Takahara T, Matsuhashi S, Yamanaka H, Minamoto T. Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. PLoS One [Internet]. 2015;10(3):e0122763. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25799582
27. Christopher L. Jerde W. Lindsay Chadderton, and David M. Lodge ARM. "Sight-unseen" detection of rare aquatic species using environmental DNA. Conserv Lett. 2011;4:150–7.
28. Takahara T, Minamoto T, Yamanaka H, Doi H, Kawabata Z. Estimation of fish biomass using environmental DNA. PLoS One [Internet]. 2012;7(4):e35868. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22563411
29. Dejean T, Valentini A, Duparc A, Pellier-Cuit S, Pompanon F, Taberlet P, et al. Persistence of environmental DNA in freshwater ecosystems. PLoS One [Internet]. 2011;6(8):e23398. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21858099
30. Willerslev PFT and E. Environmental DNA - an emerging tool in conservation for monitoring past and present biodiversity. Biol Conserv. 2015;183:4–18.
31. Caren S. Goldberg Andrew Ray, Jeremy Baumgardt, and Lisette P. Waits AS. Environmental DNA as a new method for early detection of New Zealand mudsnails (Potamopyrgus antipodarum). Freshw Sci. 2013;32(3):792–800.
32. Helen C. Rees David J. Middleditch, James R. M. Patmore, and Kevin C, Gough BCM. The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. J Appl Ecol. 2014;51:1450–9.
33. Tony Dejean Christian Miquel, Pierre Taberlet, Eva Bellmain and Claude Miaud AV. Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog Lithobates catesbeianus. J Appl Ecol. 2012;49:953–9.

34. Takahara T, Minamoto T, Doi H. Using environmental DNA to estimate the distribution of an invasive fish species in ponds. PLoS One [Internet]. 2013;8(2):e56584. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23437177
35. Christopher L. Jerde Andrew R. Mahon, Mark A. Renshaw, Joel Corush, Michelle L. Budny, Sagar Mysorekar, and David M. Lodge WLC. Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. Can J Fish Aquat Sci. 2013;70:522–6.
36. Eva Egelyng Peter Rask Moller, and Philip Francis Thomsen HC. Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. Biol Conserv. 2015;183(46–52).
37. Canada G of. Species at risk public registry: A to Z species index [Internet]. 2018. Available from: https://www.registrelep-sararegistry.gc.ca/sar/index/default_e.cfm
38. Lim MY-T, Wilson J. Thermal, morpholine and radiation stressor effects on the embryonic development of lake whitefish (Coregonus clupeaformis) and round whitefish (Prosopium cylindraceum) [Internet]. Vol. Master of, Biology. [Hamilton, ON]: McMaster University; 2016. Available from: http://hdl.handle.net/11375/20513
39. Sudbury G. Laurentian Lake [Internet]. 200 Brady Street, Sudbury, ON; 2018. Available from: https://www.greatersudbury.ca/live/environment-and-sustainability1/lake-health/fish-species-in-local-lakes/laurentian-lake/
40. O'Neill JE (Ted). Enivronmental assessment of Bruce A refurbishment for life extension and continued operations project. Ltd. GA, editor. Vol. 1, Enivronmental assessment study report. Mississauga, ON: Bruce Power; 2005.
41. Biosynthesis. TaqMan vs SBYR Green Chemistries [Internet]. 2009. Available from: http://www.biosyn.com/tew/taqman-vs-sybr-green-chemistries.aspx
42. Gail Krantzberg and C de B. A valuation of ecological services in the Great Lakes basin ecosystem to sustain healthy communities and a dynamic economy. Hamilton, ON: McMaster University; 2006.
43. Kinnunen RE. Great Lakes Commercial Fisheries. Marquette, Michigan 49855: Great Lakes Fisheries Leadership Institute; 2003.
44. Gregory L. Poe Nancy A. Conelly, Selmin Creamer, Richard C. Ready, and Richard C. Stedman BL, Unit HDR. Net benefits of recreational fishing in the Great Lakes basin: a review of the literature [Internet]. Cornell University, Ithaca, NY 14853: Department of Natural Resources; 2013. Available from: http://dnr.cornell.edu/hdru/pubs/HDRUReport13-10.pdf
45. Lupi RTM and F. Valuing recreational fishing in the Great Lakes. North Am J Fish Manag. 2013;33:1184–93.
46. Natalia V. Ivanova Robert H. Hanner and Paul D. N. Herbert TSZ. Barcoding universal primer cocktails for fish DNA barcoding. Mol Ecol. 2007;7:544–8.
47. Institute EB. Clustal Omega Multiple Sequence Alignment [Internet]. Available from: https://www.ebi.ac.uk/Tools/msa/clustalo/
48. Institute EB. T-Coffee Multiple Sequence Alignment [Internet]. Available from: https://www.ebi.ac.uk/Tools/msa/tcoffee/
49. 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;
50. Biosystems A. QuantStudio 3 and 5 Real-time PCR Systems Installation, Use, and Maintenance [Internet]. MAN0010407. 2015 [cited 2018 Nov 15]. Available from: https://assets.fishersci.com/TFS-Assets/LSG/manuals/MAN0010407_QuantStudio3_5_InstallUseMaint_UG.pdf
51. Ora L. Schlei Andrew R. Whiteley, Randy J. Brown, Jeffery B. Olsen, Louis Bernatchez and John K. Wenburg AC-L. DNA barcoding of eight North American coregonine species. Mol Ecol Resour. 2008;8:1212–8.
52. Lauren M. Overdyk Stephen S. Crawford and Robert H. Hanner HEB. Extending DNA barcoding

coverage for Lake Whitefish (Coregonus clupeaformis) across three major basins of Lake Huron. DNA Barcodes. 2015;3:59–65.
53. L. M. Overdyk A. M. Naaum, S. S. Crawford and R. H. Hanner HEB. Real-time PCR identification of lake whitefish Coregonus clupeaformis in the Laurentian Great Lakes. J Fish Biol. 2016;88:1474–640.
54. Caren S. Goldberg and David S. Pilliod KMS. Moving environmental DNA methods from concept to pracice for monitoring aquatic macroorganisms. Biol Conserv. 2015;183:1–3.
55. McCall MN, McMurray HR, Land H, Almudevar A. On non-detects in qPCR data. Bioinformatics [Internet]. 2014;30(16):2310–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24764462
56. Elvira Machler Patrick Steinmann and Florian Altermatt KD. Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. Freshw Sci. 2014;33(4):1174–83.
57. Behjati S, Tarpey PS. What is next generation sequencing? Arch Dis Child Educ Pract Ed. 2013;
58. Thomsen PF, Kielgast J, Iversen LL, Moller PR, Rasmussen M, Willerslev E. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLoS One [Internet]. 2012;7(8):e41732. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22952584
59. Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, et al. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol Ecol. 2016;
60. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol. 2012;
61. Buxton AS, Groombridge JJ, Zakaria NB, Griffiths RA. Seasonal variation in environmental DNA in relation to population size and environmental factors. Scientific Reports. 2017.

2. Smith C. 2016 Environmental Monitoring Program Report [Internet]. B-REP-07000-00009 . Bruce Pwer; 2017. Available from: http://www.brucepower.com/wp-content/uploads/2017/05/2016-EMP-Report.pdf
3. Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, Burridge M, et al. Identifying Canadian freshwater fishes through DNA barcodes. PLoS One [Internet]. 2008;3(6):e2490. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22423312
4. Ko HL, Wang YT, Chiu TS, Lee MA, Leu MY, Chang KZ, et al. Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. PLoS One [Internet]. 2013;8(1):e53451. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23382845
5. Valentini A, Pompanon F, Taberlet P. DNA barcoding for ecologists. Trends Ecol Evol [Internet]. 2009;24(2):110–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19100655
6. Teletchea F. Molecular identification methods of fish species: reassessment and possible applications. Rev Fish Biol Fish [Internet]. 2009;19(3):265. Available from: https://doi.org/10.1007/s11160-009-9107-4
7. April J, Mayden RL, Hanner RH, Bernatchez L. Genetic calibration of species diversity among North America's freshwater fishes. Proc Natl Acad Sci U S A [Internet]. 2011;108(26):10602–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21670289
8. Costa FO, Carvalho GR. The Barcode of Life Initiative: synopsis and prospective societal impacts of DNA barcoding of Fish. Genomics, Soc Policy [Internet]. 2007;3(2):29. Available from: https://doi.org/10.1186/1746-5354-3-2-29
9. Rasmussen Hellberg RS, Morrissey MT, Hanner RH. A multiplex PCR method for the identification of commercially important salmon and trout species (Oncorhynchus and Salmo) in North America. J Food Sci [Internet]. 2010;75(7):C595-606. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21535525
10. Meyer CP, Paulay G. DNA barcoding: Error rates based on comprehensive sampling. PLoS Biol. 2005;
11. Andrea Galimberti Alessia Losa, Ilaria Bruni, Silvia Federici, Maurizio Casiraghi, Stefano Martellos, Massimo Labra FDM. DNA barcoding as a new tool for food traceability. Food Res Int. 2012;50(1):55–63.

12. Benedict A. Maralit Minerva Fatimae H. Ventolero, Sweedy Kay L. Perez, Demian A. Willette, Mudjekeewis D. Santos RDA. Detection of mislabeled commercial fishery by-products in the Philippines using DNA barcodes and its implications to food traceability and safety. Food Control. 2013;33(1):119–25.
13. L. Bernatchez JJD and SB. Population bottlenecks: influence on mitochondrial DNA diversity and its effect in coregonine stock discrimination. J Fish Biol. 1989;35(Supplement A):233–44.
14. Rubinoff D, Cameron S, Will K. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. Journal of Heredity. 2006.
15. Galtier N, Nabholz B, Glemin S, Hurst GDD. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol Ecol. 2009;
16. Watanabe S, Minegishi Y, Yoshinaga T, Aoyama J, Tsukamoto K. A quick method for species identification of Japanese eel (Anguilla japonica) using real-time PCR: an onboard application for use during sampling surveys. Mar Biotechnol [Internet]. 2004;6(6):566–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15690101
17. Weller SA, Elphinstone JG, Smith NC, Boonham N, Stead DE. Detection of Ralstonia solanacearum strains with a quantitative, multiplex, real-time, fluorogenic PCR (TaqMan) assay. Appl Env Microbiol [Internet]. 2000;66(7):2853–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10877778
18. del Cerro A, Mendoza MC, Guijarro JA. Usefulness of a TaqMan-based polymerase chain reaction assay for the detection of the fish pathogen Flavobacterium psychrophilum. J Appl Microbiol [Internet]. 2002;93(1):149–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12067384
19. Kesmen Z, Gulluce A, Sahin F, Yetim H. Identification of meat species by TaqMan-based real-time PCR assay. Meat Sci [Internet]. 2009;82(4):444–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20416686
20. Keith M. Bayha Frank J. Hernandez Jr WMG. Multiplex assay to identify eggs of three fish species from the northern Gulf of Mexico, using locked nucleic acid Taqman real-time PCR probes. Aquat Biol. 2008;4:65–73.
21. Martini I. Taylor Ines Rico, Ciro Rico CF. Species-specific TaqMan probes for simultaneous identification of (Gadus morhua L.), haddock (Melanogrammus aeglefinus L.) and whiting (Merlangius merlangus L.). Mol Ecol Notes. 2002;2:599–601.
22. Trotta M, Schonhuth S, Pepe T, Cortesi ML, Puyet A, Bautista JM. Multiplex PCR method for use in real-time PCR for identification of fish fillets from grouper (Epinephelus and Mycteroperca species) and common substitute species. J Agric Food Chem [Internet]. 2005;53(6):2039–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15769133
23. Gleeson EMF and D. Environmental DNA detection of redfin perch, Perca fluviatilis. Conserv Genet Resour. 2016;8:115–8.
24. David S. Pilliod Matthew B. Laramine, Lisette P. Waits CSG. Application of environmental DNA for inventory and monitoring of aquatic species [Internet]. United States Geological Survey; 2013. Available from: https://pubs.usgs.gov/fs/2012/3146/pdf/fs2012-3146.pdf
25. Maruyama A, Nakamura K, Yamanaka H, Kondoh M, Minamoto T. The release rate of environmental DNA from juvenile and adult fish. PLoS One [Internet]. 2014;9(12):e114639. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25479160
26. Doi H, Uchii K, Takahara T, Matsuhashi S, Yamanaka H, Minamoto T. Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. PLoS One [Internet]. 2015;10(3):e0122763. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25799582
27. Christopher L. Jerde W. Lindsay Chadderton, and David M. Lodge ARM. "Sight-unseen" detection of rare aquatic species using environmental DNA. Conserv Lett. 2011;4:150–7.
28. Takahara T, Minamoto T, Yamanaka H, Doi H, Kawabata Z. Estimation of fish biomass using environmental DNA. PLoS One [Internet]. 2012;7(4):e35868. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22563411

29. Dejean T, Valentini A, Duparc A, Pellier-Cuit S, Pompanon F, Taberlet P, et al. Persistence of environmental DNA in freshwater ecosystems. PLoS One [Internet]. 2011;6(8):e23398. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21858099
30. Willerslev PFT and E. Environmental DNA - an emerging tool in conservation for monitoring past and present biodiversity. Biol Conserv. 2015;183:4–18.
31. Caren S. Goldberg Andrew Ray, Jeremy Baumgardt, and Lisette P. Waits AS. Environmental DNA as a new method for early detection of New Zealand mudsnails (Potamopyrgus antipodarum). Freshw Sci. 2013;32(3):792–800.
32. Helen C. Rees David J. Middleditch, James R. M. Patmore, and Kevin C, Gough BCM. The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. J Appl Ecol. 2014;51:1450–9.
33. Tony Dejean Christian Miquel, Pierre Taberlet, Eva Bellmain and Claude Miaud AV. Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog Lithobates catesbeianus. J Appl Ecol. 2012;49:953–9.
34. Takahara T, Minamoto T, Doi H. Using environmental DNA to estimate the distribution of an invasive fish species in ponds. PLoS One [Internet]. 2013;8(2):e56584. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23437177
35. Christopher L. Jerde Andrew R. Mahon, Mark A. Renshaw, Joel Corush, Michelle L. Budny, Sagar Mysorekar, and David M. Lodge WLC. Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. Can J Fish Aquat Sci. 2013;70:522–6.
36. Eva Egelyng Peter Rask Moller, and Philip Francis Thomsen HC. Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. Biol Conserv. 2015;183(46–52).
37. Canada G of. Species at risk public registry: A to Z species index [Internet]. 2018. Available from: https://www.registrelep-sararegistry.gc.ca/sar/index/default_e.cfm
38. Lim MY-T, Wilson J. Thermal, morpholine and radiation stressor effects on the embryonic development of lake whitefish (Coregonus clupeaformis) and round whitefish (Prosopium cylindraceum) [Internet]. Vol. Master of, Biology. [Hamilton, ON]: McMaster University; 2016. Available from: http://hdl.handle.net/11375/20513
39. Sudbury G. Laurentian Lake [Internet]. 200 Brady Street, Sudbury, ON; 2018. Available from: https://www.greatersudbury.ca/live/environment-and-sustainability1/lake-health/fish-species-in-local-lakes/laurentian-lake/
40. O'Neill JE (Ted). Enivronmental assessment of Bruce A refurbishment for life extension and continued operations project. Ltd. GA, editor. Vol. 1, Enivronmental assessment study report. Mississauga, ON: Bruce Power; 2005.
41. Biosynthesis. TaqMan vs SBYR Green Chemistries [Internet]. 2009. Available from: http://www.biosyn.com/tew/taqman-vs-sybr-green-chemistries.aspx
42. Gail Krantzberg and C de B. A valuation of ecological services in the Great Lakes basin ecosystem to sustain healthy communities and a dynamic economy. Hamilton, ON: McMaster University; 2006.
43. Kinnunen RE. Great Lakes Commercial Fisheries. Marquette, Michigan 49855: Great Lakes Fisheries Leadership Institute; 2003.
44. Gregory L. Poe Nancy A. Conelly, Selmin Creamer, Richard C. Ready, and Richard C. Stedman BL, Unit HDR. Net benefits of recreational fishing in the Great Lakes basin: a review of the literature [Internet]. Cornell University, Ithaca, NY 14853: Department of Natural Resources; 2013. Available from: http://dnr.cornell.edu/hdru/pubs/HDRUReport13-10.pdf
45. Lupi RTM and F. Valuing recreational fishing in the Great Lakes. North Am J Fish Manag. 2013;33:1184–93.
46. Natalia V. Ivanova Robert H. Hanner and Paul D. N. Herbert TSZ. Barcoding universal primer cocktails for fish DNA barcoding. Mol Ecol. 2007;7:544–8.
47. Ora L. Schlei Andrew R. Whiteley, Randy J. Brown, Jeffery B. Olsen, Louis Bernatchez and John

K. Wenburg AC-L. DNA barcoding of eight North American coregonine species. Mol Ecol Resour. 2008;8:1212–8.
48. Lauren M. Overdyk Stephen S. Crawford and Robert H. Hanner HEB. Extending DNA barcoding coverage for Lake Whitefish (Coregonus clupeaformis) across three major basins of Lake Huron. DNA Barcodes. 2015;3:59–65.
49. L. M. Overdyk A. M. Naaum, S. S. Crawford and R. H. Hanner HEB. Real-time PCR identification of lake whitefish Coregonus clupeaformis in the Laurentian Great Lakes. J Fish Biol. 2016;88:1474–640.
50. Caren S. Goldberg and David S. Pilliod KMS. Moving environmental DNA methods from concept to pracice for monitoring aquatic macroorganisms. Biol Conserv. 2015;183:1–3.
51. McCall MN, McMurray HR, Land H, Almudevar A. On non-detects in qPCR data. Bioinformatics [Internet]. 2014;30(16):2310–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24764462
52. Elvira Machler Patrick Steinmann and Florian Altermatt KD. Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. Freshw Sci. 2014;33(4):1174–83.
53. Behjati S, Tarpey PS. What is next generation sequencing? Arch Dis Child Educ Pract Ed. 2013;
54. Thomsen PF, Kielgast J, Iversen LL, Moller PR, Rasmussen M, Willerslev E. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLoS One [Internet]. 2012;7(8):e41732. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22952584
55. Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, et al. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol Ecol. 2016;
56. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol. 2012;
57. Buxton AS, Groombridge JJ, Zakaria NB, Griffiths RA. Seasonal variation in environmental DNA in relation to population size and environmental factors. Scientific Reports. 2017.

Appendix A: DNA barcoding vs. morphological identification of larval fish and embryos in Lake Huron: advantages to a molecular approach

DNA barcoding vs. morphological identification of larval fish and embryos in Lake Huron: advantages to a molecular approach

Emily N. Hulley; Natalie D. J. Taylor, Andrew M. Zarnke, Christopher M. Somers, Richard G. Manzon, Joanna Y. Wilson and Douglas R. Boreham

(Accepted for publication in the Journal of Great Lakes Research)

Abstract

The Great Lakes provide habitat to over 160 species of freshwater fish, many of which are ecologically and economically important. Concern for management and conservation of declining fish populations makes it important that accurate identification techniques are used for environmental monitoring programs. DNA barcoding may be an effective alternative to morphological identification for industrial monitoring programs of larval and embryonic fish, but comparisons of the two approaches with species from the Great Lakes are limited. It may be particularly important to examine this issue in the Great Lakes because a relatively young group of post-glacial fish species are present, which may be difficult to resolve using morphology or genetics. Six hundred and fifty seven larval fish were identified from Lake Huron (Ontario, Canada), using morphology and DNA barcoding. DNA barcoding was used to identify 103 embryos that morphology could not identify. Morphological identification and DNA barcoding had a percent similarity of 76.9%, 96.6% and 96.6% at the species, genus, and family levels, respectively. Thirty-seven specimens were damaged and unidentifiable using morphology; 35 of these were successfully identified using DNA barcoding. However, 23 other specimens produced no PCR product for barcoding using 2 different primer sets. Discrepancies between morphology and DNA barcoding were driven by 2 major factors: inability of COI to resolve members of the genus *Coregonus*; limited resolution of morphological features for *Catostomus* and Cyprinidae. Both methods have pros and cons; however, DNA barcoding is more cost-effective and efficient for industrial monitoring programs.

Keywords*:*

DNA barcoding, morphological identification, larval fish, cytochrome oxidase I, fish embryo

Introduction

Industrial water usage, such as once-through cooling, may impact aquatic species across all life stages. Once-through cooling systems brings in lake water which is pumped into condenser units for cooling, then is released as a warmed surface effluent back into the lake. The effluents may contain trace amounts of chemical or radiological contaminants, exposing fish in the near-shore area, and the intake process may result in the impingement and entrainment of aquatic organisms (Bruce Power 2005). Environmental monitoring programs monitor fish that are impinged at intake points and attempt to identify the species and numbers affected (Ko *et al.* 2013; Moura *et al.* 2008; Valdez-Moreno *et al.* 2010; Richardson *et al.* 2007). However, many of the individuals captured by water intakes are in the larval stage, this can be very difficult to identify based on morphology. Several species share very similar traits at the larval stage, which makes accurate identification difficult without highly specialized training (Ko *et al.* 2013). In addition, in some cases the larval stage may appear quite different among members of the same species based on relative age (Teletchea 2009; Strauss and Bond 1990). Difficulties with identification can be compounded when larval fish specimens are damaged while travelling through the cooling system (Teletchea 2009; Strauss and Bond 1990). As well, morphological identification is generally unable to identify fish in the embryo stage due to the lack of morphological indicators (Teletchea 2009; Strauss and Bond 1990). As a result, any attempt at morphological identification requires highly trained taxonomists, and even with this training, it is recommended that larval fish only be identified to the family level (Ko *et al.* 2013). However, ecological studies and environmental monitoring often require information at the genus or species level, so better resolution is required.

It has been shown that the mitochondrial cytochrome oxidase I (COI) gene can serve as the universal genetic barcode to identify many different organisms (Hubert *et al.* 2008). Over the last decade, DNA barcoding has become a well-established technique used in a variety of different settings, from identifying fish market substitution (Barbuto *et al.* 2010; Galimberti *et al.* 2013; Pinto *et al.* 2013; Wong and Hanner 2008) to species monitoring and conservation (Ardura *et al.* 2010; Hajibabae *et al.* 2007; Valdez-Moreno *et al.* 2012; Ward *et al.* 2008). Ko *et al.* (2013) examined DNA barcoding to evaluate the accuracy of

traditional morphological identification of Taiwan larval fishes by five different larval fish taxonomists. Marine larval fish were collected in the northern, southern and north-western seawaters of Taiwan over a two-year period yielding a sample size of 100 specimens per year. From the 100 samples, 12 samples failed PCR and could not be identified; the remaining 88 were DNA barcoded to 87 families, 79 genera and 69 species. Each of the five taxonomists from separate laboratories then identified the species to the three taxanomic categories. Ko *et al.* (2013) found that the consistency of identification between the five laboratories was very low: ~80.1% for family level, ~41.1% for genus and ~13.5% for species level. Consequently, Ko *et al.* (2013) proposed that morphological identification of larval fish should be more conservative and that DNA barcoding is a useful technique to verify the accuracy of larval identification between different taxonomists. However, few studies have addressed similar issues in other systems with different fish communities.

The Great Lakes specifically provide habitat to over 160 species of freshwater fish. These fish are not only ecologically important for their surrounding ecosystem but generate revenue in excess of 8 billion dollars annually (Krantzberg 2006). Many of these freshwater fish are post-glacial, recently diverged, and exhibit interspecies COI haplotype sharing (April *et al.* 2011; Hubert *et al.*2008; Kochzius *et al.* 2010; Sajdak and Phillips 1995; Renaut *et al.* 2009). For example, the genus *Coregonus* contains over 30 freshwater species, many of which have both ecological and economical importance in the Great Lakes (Schlei *et al.* 2008). Furthermore, several of these species are listed as at risk of extinction, including shortnose cisco (*C. reighardi*), Great Lakes kiyi (*C. kiyi*) and shortjaw cisco (*C. zenithicus*) (Ontario 2017). The *Coregonus* genus has had a relatively recent evolutionary divergence making it very difficult to identify species morphologically, especially when in their larval stages (Schlei *et al.* 2008). Coregonid larvae are morphologically similar and may also be more genetically alike than the marine species identified by Ko *et al.* (2013). Thus, validation of molecular approaches for larval freshwater species in the Great Lakes is critical.

Here we compare the morphological identification of larval fish and embryos collected from the water intake system at a large nuclear power facility on Lake Huron, with DNA barcoding of the same

specimens (Thome *et al.* 2016, Graham *et al.* 2016). Overall, our objective was to establish which method of identification was more accurate and cost effective for long-term environmental monitoring programs for large industrial operations using once-through cooling systems in the Great Lakes. We were also specifically interested in the accuracy of morphological identifications of lake whitefish (*C. clupeaformis*) and deepwater sculpin (*Myoxocephalus thompsonii*). Lake whitefish is ecologically and culturally important in Lake Huron, supporting a large commercial fishery and First Nations subsistence fishery (O'Neill 2005; Overdyk *et al.* 2015; Schlei *et al.* 2008). Deepwater sculpin is listed as a species of special concern because of declining populations in the Laurentian Great Lakes (COSEWIC 2015). Our work will help guide appropriate monitoring of industrial impacts on larval fish in the Great Lakes.

Methods

Sample Collection

Bruce Power is a nuclear power plant located on the shores of Lake Huron in Tiverton, Ontario. The plant consists of 8 CANDU pressurized heavy water reactors across two stations (4 reactors in Bruce A and 4 reactors in Bruce B). Both stations rely on once-through cooling for steam condensation and have their own condenser cooling water circuit. Water is collected from Lake Huron via an intake tunnel located offshore on the lake bottom. Water travels through the intake tunnel into the forebay, which is located on shore adjacent to the Bruce A and B stations. From the forebay, water is pumped into the turbine hall for steam condensation and is subsequently released back into the lake via a surface discharge channel (Bruce Power 2015). Fish larvae and embryo specimens were collected from March through December in 2013 and 2014 within the Bruce A forebay. Samples were collected using a round, 500µm mesh plankton net that was 0.72m in diameter and was lowered 3m into the intake water for 5 to 140 minutes (median 36 minutes). Specimen collection occurred both during the day and at night. Between 2013 and 2014, there were 81 day and 80 night sampling efforts, with a minimum of 3 nets set per effort. Sampling was done during the day and night to avoid any bias due to nocturnal or diurnal behaviour of any specific species. Each collected specimen was given a unique identification number and larval fish

were stored in 95% ethanol until morphological analysis. After morphological analysis, larval fish of the same species from the same collection time were stored together in 95% ethanol. Fish embryos from the same collections were stored together in 95% ethanol. All efforts resulted in the collection of 1740 larval fishes and 2831 fish embryos. Larval specimens were randomly chosen for both morphological identification and DNA barcoding so as to capture a representative sample from all possible species. Fish embryos were randomly selected for DNA barcoding analysis.

Morphological Identification

Samples were analyzed commercially (through an external contractor) by an expert ichthyologist who specializes in the identification of larval fishes from the Laurentian Great Lakes. Specimens were identified based on body shape, myomeres, pigmentation, meristic count, and fin characteristics (e.g., number, shape, relative position etc.). When possible, specimens were identified to the species level; otherwise, specimens were identified to the genus or family level. Results of morphologic identifications were recorded based on the unique identification number of each specimen. Fish eggs were not identified morphologically. The cost of the larval identifications was recorded for a comparison with DNA barcoding. Cost was calculated on a per-specimen basis in US Dollars (USD) and converted to Canadian Dollars (CAD).

Molecular Identification

DNA was extracted from individual larval fish and embryos using spin column kits according to manufacturer guidelines (Qiagen DNEasy, Mississauga, ON; Norgen Biotech DNA extraction kit, Thorold, ON). When larval specimens were small (<12mm in total length), the entire fish was used for DNA extraction; when larval specimens were larger, a portion of the body (up to 12mm) was used for DNA extraction. DNA concentration from extractions was measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific) or Qubit fluorometric quantation (Life Technologies). All 260/280 and 260/230 ratios were within appropriate range of 1.8-2 for DNA analysis.

For all specimens, a 658 bp region of the COI mitochondrial genome was PCR-amplified using universal primers FishF1 (5'-TCA ACC AAC CAC AAA GAC ATT GCC AC-3') and FishR1 (5'-TAG ACT TCT GGG TGG CCA AAG AAT CA-3') (Ward *et al.* 2005). PCR reactions consisted of a total volume of 25µL with components as follows: 1x PCR buffer; 2.5mM of $MgCl_2$; 0.1µM of each forward and reverse primers, 0.05mM of each dNTPs, 0.31 units of *Taq* DNA polymerase, and 10ng of template DNA. The thermal cycling regime consisted of: 2 minutes at 94°C followed by 35 cycles of: 30 seconds at 94°C, 40 seconds at 52°C and 1 minute at 72°C. Final extension was for 10 minutes at 72°C. PCR products were separated on a 1% agarose gel to verify the presence of a product in the target size range. Specimens that failed the initial PCR were run a second time using the universal fish primers FF2d (5'-TTC TCC ACC AAC CAC AAR GAY ATY GG-3') and FR1d (5'-CAC CTC AGG GTG TCC GAA RAA YCA RAA-3') (Ivanova *et al.* 2007) using the reaction conditions specified above. Negative control samples (with no template) were randomly dispersed throughout the PCR runs to ensure that there was no contamination.

Successfully amplified DNA was purified using the Qiagen MiniElute PCR Purification Kit (Mississauga, Ontario) with a final elution volume of 10µL. Final elution was performed using MilliQ biology-grade water or nuclease-free water; 3.2pmol of forward primer (FISH F1 or FF2d) was added to the purified product and samples were Sanger sequenced (University of Calgary, Core DNA Services). The overall cost for the identification of the specimens using DNA barcoding was recorded for later comparison. Cost was calculated on a per specimen basis, in CAD and included cost of labour, reagents used, sequencing and analysis. All processes adhered to a mainstream pipeline in the laboratory to ensure methods were standardized and easily transferable to non-specialized facilities and personnel.

Data Analysis

Sequences were aligned and compared using SeaView V. 4.5.4 (Gouy *et al.* 2010). Specimens were identified by comparing COI sequences obtained against those in the NCBI database using the Basic Local Alignment Search Tool (BLAST) (https://blast.ncbi.nlm.nih.gov/). We used quantitative criteria

similar to those described by Ko *et al.* (2013) for identifications. Specifically, specimens were identified to the species level when they matched an individual species at >98% similarity and the maximum bit score returned by BLAST was over 1000, and with the bit score of the next most likely species match lower by more than 100 bit points. When a clear top match was not present, i.e., when there was more than one species with >98% match, or no species >98% match, samples were identified to the top-matching genus or family. Specimens were identified to the genus and family when the similarity values were 97%-92% and 91%-84%, respectively. Geographic distribution of fish species was taken into account when identifying the fish specimens; i.e., fish species that do not occur in the study region were not included for consideration.

COI sequences for each of the identified species were collected from the Barcode of Life Database (BOLD) (http://www.boldsystems.org/index.php/). Sequences were aligned using Clustal Omega-EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/) to perform a percent identity matrix for species similarity.

Results

Morphological Identification

Of the 1740 larval fishes and 2831 fish embryos sampled, 657 larval fish and 103 embryos were randomly selected for this study. From the starting 657 larval fish, 620 were identified to 9 families (94.4%), 618 to 13 genera (94.1%), and 583 to 18 species (88.7%). Of the total specimens analysed, 37 (5.6%) could not be identified because they were damaged in some way (i.e., missing key features for identification). Morphological identification indicated that species exceeding 5% of our collections were: burbot (*Lota lota*), deepwater sculpin, bloater (*C. hoyi*), rainbow smelt (*Osmerus mordax*), and round goby (*Neogobius melanostomus*) (Table 1). Burbot and deepwater sculpin were at least two-times more abundant than the next closest species (bloater). Morphological identification was unable to identify any of the embryos.

Table 1 - Total number of larval fishes that were both morphologically and molecularly. identified.

Family	Genus	Species	# of fish identified using morphology (%)	# of fish identified using DNA barcoding (%)
Clupeidae	***Alosa***	***pseudoharengus***[1]	1 (0.2)	1 (0.2)
Catostomidae	***Carpoides***	***cyprinus***[2]	2 (0.3)	3 (0.5)
	Catostomus	**Genus only**	20 (3.0)	0
		commersonii[3]	0	17 (2.6)
		catostomus[4]	0	1 (0.2)
Salmonidae	***Coregonus***	**Genus only**	15 (2.3)	100 (15.2)
		artedi[5]	27 (4.0)	0
		hoyi[6]	60 (9.1)	0
		clupeaformis[7]	7 (1.0)	4 (0.6)
	Oncorhynchus	***tshawytscha***[8]	1 (0.2)	0
		mykiss[9]	0	1 (0.2)
Cyprinidae	**Family only**	**Family only**	2 (0.3)	0
	Notropis	***atherinoides***[10]	0	5 (0.8)
		hudsonius[11]	0	1 (0.2)
	Cyprinus	***carpio***[12]	0	1 (0.2)
Gadidae	***Lota***	***lota***[13]	243 (37.0)	257 (39.1)
Cottidae	***Myoxocephalus***	***thompsonii***[14]	144 (21.9)	147 (22.4)
Gobiidae	***Neogobius***	***melanostomus***[15]	43 (6.5)	44 (6.7)
Osmeridae	***Osmerus***	***mordax***[16]	48 (7.3)	43 (6.5)
Percidae	***Perca***	***flavescens***[17]	6 (0.9)	6 (0.9)
	Sander	***vitreus***[18]	1 (0.2)	1 (0.2)
		Unidentified	37 (5.6)	25 (3.8)
		Total	657	657

Common Fish Names: 1 – Alewife, 2 – Quillback, 3 – White Sucker, 4 – Longnose Sucker, 5 – Cisco, 6 – Bloater, 7 – Lake Whitefish, 8 – Chinook Salmon, 9 – Rainbow Trout, 10 – Emerald Shiner, 11 – Spottail Shiner, 12 – Common Carp, 13 – Burbot, 14 – Deepwater Sculpin, 15 – Round Goby, 16 – Rainbow Smelt, 17 – Yellow Perch, 18 – Walleye

Molecular Identification

Of the 657 larval fish specimens analysed using COI sequence data, 632 were identified to 9 families (96.1%), 632 to 13 genera (96.1%), and 532 to 18 species (81.0%). Twenty-five (3.8%)

specimens could not be identified due to amplification failure with both sets of PCR primers, 23 of which were identified using morphology. However, of the 35 of the 37 damaged specimens that could not be identified morphologically were identified to the species level using DNA barcoding.

DNA barcoding was able to identify lake whitefish but could not differentiate between cisco (*C. artedi*), kiyi, and bloater. The 100 specimens (17% of sample population) that were cisco, kiyi or bloater were therefore only identified to the genus. According to DNA barcoding, species exceeding 5% of our collections were: burbot, deepwater sculpin, the genus *Coregonus* (excluding lake whitefish), rainbow smelt, and round goby (Table 1).

We were able to amplify a PCR product from 52 (50.5%) of the 103 individual fish embryos collected. All 52 (100%) of the embryos identified were walleye (*Sander vitreus*). These eggs were from 14 different collection efforts that took place in April (2 collections), May (5 collections) and June (2 collections) in 2013, and in May (4 collections) and June (1 collection) in 2014. COI data for walleye embryos contained several different haplotypes across the sampling periods, suggesting that embryo intake is common and includes a variety of females.

Species Similarity Analysis

All species shared a percent similarity of 76.99 or greater with one another (Table 2). *Coregonus*, *Catastomus*, *Notropis* and *Oncorhynchus* all demonstrated the highest similarity between their genera species, with percent similarity of 99.69 and 97.85, 91.10, 90.95 and 93.56 respectively. Cisco and bloater exhibited the greatest similarity between their COI sequences out of all the identified species with a percent similarity of 99.69% (Table 2).

Table 2. COI gene sequence similarity analysis of the 18 different species identified in the present study. Percent similarity is represented as the number of base pair matches in the COI sequences between the study species. Corresponding codes for the species are listed below the matrix.

	DWS	RB	BURB	Al	ES	SS	CC	Ql	WS	LS	RS	YP	W	Ci	Bl	LWF	CS	RT
DWS	100.0																	
RG	81.3	100.0																
BURB	80.4	79.9	100.0															
Al	79.6	78.7	80.4	100.0														
ES	78.5	77.9	79.8	80.4	100.0													
SS	77.0	78.1	79.9	80.1	91.0	100.0												
CC	79.3	78.5	79.6	81.4	84.4	84.8	100.0											
Ql	79.1	78.4	80.5	82.5	84.5	83.3	86.7	100.0										
WS	78.5	78.7	78.8	79.3	83.1	84.5	85.3	86.5	100.0									
LS	77.9	79.0	80.2	81.3	84.4	85.1	85.0	87.6	91.1	100.0								
RS	79.1	78.8	78.1	79.1	78.5	78.4	77.2	79.3	77.6	77.5	100.0							
YP	79.8	80.5	79.0	78.2	77.9	78.8	80.2	77.6	78.7	77.6	78.5	100.0						
W	80.5	80.4	78.8	79.6	78.5	78.8	78.7	79.0	77.9	77.9	79.6	85.1	100.0					
C	79.0	78.7	79.0	79.6	77.5	78.2	77.9	78.7	77.9	78.8	79.1	80.7	79.8	100.0				
Bl	79.0	78.7	79.0	79.6	77.2	77.9	77.9	77.7	77.9	78.8	79.1	81.0	80.1	99.7	100.0			
LWF	78.8	78.7	78.8	80.7	77.2	77.6	78.4	79.1	77.8	78.8	79.6	81.1	79.8	97.9	97.9	100.0		
CS	79.3	80.4	78.8	79.5	78.7	80.1	79.0	78.1	77.9	78.5	78.1	80.7	80.1	84.5	84.4	84.5	100.0	
RT	79.1	79.1	78.1	79.1	77.3	79.6	79.0	78.1	78.7	78.8	78.1	80.7	79.9	85.1	85.0	85.1	93.6	100.0

DWS - Deepwater Sculpin, RB - Round Goby, BURB - Burbot, Al - Alewife, ES - Emerald Shiner, SS - Spottail Shiner, CC - Common Carp, Ql - Quillback, WS - White Sucker, LS - Longnose Sucker, RS - Rainbow Smelt, YP - Yellow Perch, W - Walleye, C - Cisco, Bl - Bloater, LWF - Lake Whitefish, CS - Chinook Salmon, RT - Rainbow Trout

Morphological Identification vs. DNA Barcoding

Morphological and molecular identification techniques demonstrated 3 major differences with respect to identifying and differentiating the Lake Huron larval fishes. The first was with differentiating members of the *Coregonus* genus. The expert ichthyologist identified 15 (2.3%) to genus level, whereas DNA barcoding differentiated 100 (16.8%) to genus level (Table 1). The second major difference was observed in *Catostoma* and Cyprinidae. Morphology was limited to differentiating individuals to the genus or family level, whereas DNA barcoding was capable of identification to the species level. This difference between the techniques represented 25 (4.0%) misidentifications at the species level (Table 1). Lastly, the other major discrepancy between techniques occurred at the species level for lake whitefish. Morphology identified 7 individuals as lake whitefish, but DNA barcoding identified only 4 (Table 1). Molecular identification also revealed that other *Coregonus* species, specifically bloater, were morphologically identified as larval lake whitefish (Table 3). The remaining discrepancy was that DNA barcoding identified a specimen as a lake whitefish, which morphology had identified as a member of the *Catostomus* genus (Table 3). Both *Coregonus* and *Catastomus* had one of the greatest percent homology between its species, with percent homologies of 99.69 and 97.85 for *Coregonus* and 91.10 for *Catastomus* species (Table 2). Discrepancies for rainbow smelt were also observed; 5 of 48 specimens morphologically identified as rainbow smelt were identified as burbot, emerald shiner (*N. antherinoides*), or yellow perch (*Perca flavescens*) with DNA barcoding (Table 3). Two specimens were morphologically identified as burbot, but DNA barcoding identified the species as rainbow smelt. In contrast, only 3 of 144 deepwater sculpin identifications disagreed; 1 specimen was morphologically identified as a bloater and 2 as burbot. In addition, morphological identification was unable to identify any of the embryo samples in this study. DNA barcoding however was able to distinguish 52 (50.5%) of the 103 individual fish embryos collected as walleye.

Table 3– Comparison of morphologic and genetic identification with respect to number of unidentified specimens. Genetic identification was used to verify the accuracy of morphology identifications; species found to be misidentified are reported below. Individuals that were unidentified by morphology and/or DNA barcoding are indicated in the last two columns.

Species	**Total Number of Samples Analyzed**	**Morphology Misidentification (%)**	**Morphology Unidentified (%)**	**Genetic Unidentified (%)**
Alewife (*Alosa pseudoharengus*)	1			
Quilback (*Carpiodes cyprinus*)	3			
Sucker (*Catostomus spp.*)		2 (0.3)		
Longnose Sucker (*C. catostomus*)	17			
White Sucker (*C. commersonii*)	1			
Whitefish (*Coregonus spp.*)	100	2 (0.3)	2 (0.3)	
Cisco (*C. artedi*)				4 (0.6)
Lake Whitefish (*C. clupeaformis*)	4	4 (0.6)		
Bloater (*C. hoyi*)		4 (0.6)		
Carp/Minnow (*Cyprinidae*)		2 (0.3)		
Common Carp (*Cyprinus carpio*)				
Burbot (*Lota lota*)	257	7 (1.1)	27 (4.1)	12 (1.8)
Deepwater Sculpin (*Myoxocephalus thompsonii*)	147		4 (0.6)	4 (0.6)
Round Goby (*Neogobius melanotstomus*)	44			
Rainbow Trout (*Oncorhynchus mykiss*)	1			
Chinook Salmon (*Oncorhynchus tshawytscha*)		1 (0.2)		
Rainbow Smelt (*Osmerus mordax*)	43	5 (0.8)		3 (0.5)
Yellow Perch (*Perca flavescens*)	6			
Emerald Shiner (*Notropis atherinoides*)	5		2 (0.3)	
Spottail Shiner (*Notropis hudsonius*)	1			
Walleye (*Sander vitreus*)	1			
Unidentified	26		2 (0.3)	3 (0.5)
Total	**657**	**27**	**37**	**26**

In terms of cost, the external contractor charged $10 USD to identify each larval fish (~$13.69 CAD), while DNA barcoding cost $10.18 CAD per specimen (Table 4), leading to a difference in cost of $3.12 per larval fish. In terms of time, both techniques required similar amounts of labour time by an individual. The external contractor spent roughly 8 to 16 hours to identify about a hundred samples while human preparation for DNA barcoding before running on a machine was approximately 4 hours (Table 4). Variation in time with morphological identification was a result of species similarity; species from the *Coregonus* genus require more time by a highly experienced taxonomist to identify, whereas burbot can be quickly identified by a laboratory technician. The time required for molecular identification was dependent on the quality and type of sample.

Table 4 - The breakdown of cost and time for the genetic and morphological identification of larval fishes in Canadian dollars.

	Cost per Specimen	**Total Cost**	**Time per 96 Samples**
DNA Barcoding			
QiagenDNEasy Blood and Tissue Kit	$3.30	$2,168.10	Prep: 1 hour Incubation: 3-4 hours
QiagenTaq PCR Core Kit	$0.80	$525.60	Prep: 30 minutes Run Time: 1.5 hours
Primers	$0.03	$19.71	/
Agarose gel and Loading Dye	$0.18	$118.26	Prep: 30 minutes Run Time: 1-2 hours
QiagenMinElute PCR Purification Kit	$2.61	$1,714.77	1 hour
Sequencing	$3.50	$2,299.50	/
Analysis	/	/	1 hour
Labour	$1.04	$684.38	/
Total	**$11.46**	**$7, 529.22**	**Prep: 4 hours* Run Time: 5.5-8.5 hours***
Morphological Identification			
Labour	$10 USD ($13.69CAD)	$6,750USD ($8,9994.33CAD)	8-16 hours
Total	**$13.69**	**$8,9994.33**	**8-16 hours****

*With prep representing the length of time each sample requires human preparation before running on a machine.
**Morphological identification time varies depending on species complexity, species such as *Coregonus* would require significantly more time and experience to identify than a burbot.

Discussion

In the work presented here, we compared the precision, cost and efficiency of morphological and DNA barcoding techniques with the identification of 632 larval fish specimens and 103 embryos from Lake Huron. DNA barcoding proved to be the more suitable choice of technique for large-scale industrial environmental monitoring for freshwater larval fish in the Great Lakes due to its accurate, cost-effective, and efficient nature, specifically with species of concern and economical importance.

Species of Interest

The accurate identification of deepwater sculpin (species of special concern) (COSEWIC 2015) and lake whitefish (commercial fishery and First Nations subsistence fishery) (O'Neill 2005; Overdyk *et al.* 2015; Schlei *et al.* 2008) were of particular interest for this study. For deepwater sculpin only a 2%

discrepancy rate was found between the two techniques. A possible reason for this close level of agreement between the two techniques is the lack of closely related and therefore morphologically similar species in Lake Huron that could lead to misidentification (Blood and Matarese 2010; Richardson and Washington 1980; Khan and Faber 1974). In contrast, lake whitefish were misidentified at a higher rate using morphology. Many of the discrepancies found included morphological identification of a specimen as a lake whitefish, but DNA barcoding identified the same specimen as a longnose sucker. A similar error was also fairly common involving cisco and white sucker. Misidentifications are important to avoid with respect to environmental monitoring of industrial operations as they can lead to inappropriate measures to reduce ecological impacts that may not actually exist. The equivocal performance of the two techniques for deepwater sculpin and the potential for inaccurate identification of lake whitefish using morphology suggest that DNA barcoding is the more accurate and effective method to identify these two species in an industrial setting.

Disadvantages to the Morphological Approach

For the Cyprinidae family, morphological features were unable to identify any of the specimens beyond the family, while DNA barcoding was able to identify these specimens to the species. Morphology also frequently misassigned members of this family to others, such as the *Myoxocephalus* and *Neogobius* genus. DNA barcoding was able to differentiate the majority of species the Cyprinidae family from one another due to the high level of interspecific variability in the COI region. April et al. (2011) found that of the 221 species they analyzed from the Cyprinidae family, only 10 species could not be differentiated using DNA barcoding. The ichthyologist was likely unable to identify these specimens beyond the family level as there are defined criteria for their delineation, as well as the specimens could have been damaged and missing key identification features (Fuiman *et al.* 1983).

In terms of the *Catostomus* genus, DNA barcoding clearly identified specimens to lower taxonomic units than using morphology, which only identified these specimens to the genus while DNA barcoding resolved them to the species. DNA barcoding was able to easily differentiate longnose sucker

(*Catostomus catostomus*) from white sucker (*C. commersonii*) due to the variability in their COI sequences. Our analysis showed that there were 52 consistently variable base pairs between these two species, making their differentiation using DNA barcoding clear and simple. Contrarily, there was great difficulty to resolve these species beyond the genus level using morphology. This is not a difficulty that has been limited to this study. It has been noted that morphology cannot differentiate longnose and white sucker until scales have formed, which occurs during the juvenile stage (Synder 1998; Kay *et al.* 1994). This difficulty can partially be attributed to the lack of difference in their pigmentation and the fact that their myomere numbers are very similar (Fuiman and Witman 1979).

Disadvantages to DNA Barcoding

A major problem for potential industrial applications of DNA barcoding with the COI gene is that it is unable to differentiate bloater, kiyi, and cisco. These three species share very little variation in their COI sequences, scoring percent homologies of 99.69 for cisco and bloater, 97.85 for lake whitefish with bloater and cisco (Table 2). Due to extremely high percent homology, these specimens in our collection could only be identified to the genus level using the molecular approach. This is a persistent problem with the *Coregonus* genus (Schlei *et al.* 2008; April *et al.* 2011; Hubert *et al.* 2008). April et al. (2011) showed that there were 7 species in the *Coregonus* genus that were indistinguishable when DNA barcoding was used (including bloater, kiyi and cisco). Bloater, kiyi and cisco specimens comprised 17% of the collection we assessed, so members of the *Coregonus* genus had a major influence on our perception of the performance of DNA barcoding. However, there are no means of confirming the precision of morphological identification of bloater and cisco, so we cannot truly evaluate the performance of the two techniques for this difficult group. It is possible that DNA barcoding cannot differentiate between these three species due to hybridization or recent divergence (Renaut *et al.* 2009; Nolte *et al.* 2009).

Identification of Fish Embryos

We were able to generate DNA barcode data from individual fish embryos in over half of the collected samples. Challenges with embryo identification using barcoding could include poor DNA

quality and/or low DNA concentrations resulting in little to no amplification of product. All of the embryos that were barcoded were identified as walleye, which is a valuable commercial and recreational species. Multiple spawning females could have deposited embryos in an area of the lake near the water intake system, or been physically impinged. This information may allow for additional considerations for walleye at this location. Because of the challenges with other techniques, molecular identification of fish embryos will provide industries that utilize lake, river and seawater with a more accurate picture of fish species impacted at the spawning or embryo development stages.

Advantages to DNA Barcoding

Another factor to consider is the accuracy of both techniques in terms of identification to the species level. Morphological misidentifications to the species level occurred 3.5% of the time, a total of 21 individuals. The success of morphological identifications depends on the life-stage and quality of the specimen, while the success of DNA barcoding depends on the degree of interspecific variation at the CO I locus [10, 17]. DNA barcoding has been shown to be 99-100% accurate when a comprehensive database for comparison is available. With respect to fish there are extensive publically available datasets for such comparison (Ko *et al.* 2013; Meyer and Paulay 2005). Our study shows that lake whitefish, rainbow smelt, and burbot are the most likely to be misidentified when identified to the species level using morphology (71%, 16%, and 5% misidentification rates respectively). The difficulty associated with morphological identification to the species can have serious consequences on study results. Morphology often identified specimens as lake whitefish, rainbow smelt and burbot, but these specimens were then found to be another species through DNA barcoding (Table 3). These misidentifications can lead to an overestimation of the ecological impact on these species, and an underestimation of the impact on other species. For environmental studies, these misidentifications can lead to erroneous repopulation efforts, and erroneous controls and quotas on these species, all of which can result in the mismanagement of the fish population. Morphological identification by an expert ichthyologist can be highly accurate, however DNA barcoding demonstrates more accurate results, specifically with an industrial application.

DNA barcoding was able to identify more specimens (96.2%) than morphological identification (94.4%) when all taxonomic levels are considered. This is because DNA barcoding does not rely on the physical quality of the specimen appearance and can accurately identify larval fish independent of the life-stage (Ko *et al.* 2013). As such, barcoding is able to identify specimens that have been damaged, are missing key diagnostic features, or that are in stages of development that cannot be identified using morphology (Ko *et al.* 2013). All larval fish specimens that could not be identified morphologically were damaged in some way; e.g., some specimens were not complete and others were not preserved appropriately. DNA barcoding was able to identify 95% of these specimens, demonstrating the versatility of DNA barcoding. As such, DNA barcoding should be the technique used when the specimen is damaged in any way. However, when the specimen is not damaged, both morphological identification by an expert ichthyologist and DNA barcoding are robust means of larval fish identification.

In addition to accuracy, industrial level environmental monitoring programs rely on highly efficient and cost-effective methods to identify impacted individuals. Thousands of individuals are impinged in once-through cooling systems each year. In 2013, 25,000 individuals were impinged compared to the following where 16,000 were impinged from the same system (Smith 2016). DNA barcoding provided a cheaper and more time efficient for identifying larval fishes than morphological identification (Table 4). The external contractor charged $10 USD to identify each larval fish (~$13.69 CAD), whereas DNA barcoding, with a technician (salary of $20 per hour), cost $11.46 CAD per specimen leading to a difference in cost of $2.23 per larval fish. It is important to note that both techniques have additional costs that are not listed in Table 3. For DNA barcoding these include additional start-up costs of approximately $20 to $25,000 (CAD) to create a laboratory that is capable of preparing DNA for Sanger sequencing. Additional equipment required would include the purchase and maintenance of a PCR thermal cycler, centrifuge, water bath and spectrophotometer. For morphological identification, there are additional costs for start up equipment as well; however these are substantially cheaper ($1 to $2,000 dollars (CAD)) as a microscope and light source would be the only required equipment. In terms of specialized training and human preparation needed to process samples, DNA

barcoding presents a more advantageous approach. Both techniques had similar lengths of time to process a hundred samples (Table 4). However, DNA barcoding only required approximately 4 hours of preparation by a technician before it was run on a machine, compared to the 8 to 16 hours spent by a taxonomist morphologically identifying samples under a microscope. DNA barcoding time could be further reduced by eliminating the ethidium bromide staining and gel electrophoresis step for a more high-throughput system. This would decrease the total processing time to 3.5-6.5 hours. DNA barcoding also requires cheaper and less specialized training to identify larval and embryo fishes compared to morphological identification. To become an expert taxonomist it takes years to decades of specialized training and courses. Preparing samples for Sanger sequencing can be prepared by laboratory technician (two year college degree) with minimal additional training. Finally, damaged specimens can still provide useful DNA for barcoding, even when they are missing all identifying morphological features. DNA barcoding also provides a chance to identify embryo specimens that morphological identification cannot. Thus, we conclude that although the performance of the two techniques is very similar in terms of accuracy, DNA barcoding offers distinct advantages that make it a more advantageous option for large scale industrial environmental monitoring programs.

Conclusions

DNA barcoding and morphological identification provide equivocal performance overall in terms of resolution. However, DNA barcoding has the advantage of being more cost-effective, more efficient, and requiring significantly less training than morphological identification. The procedure could be modified for high throughput analyses, further reducing the cost and time required for the identification of larval fish and fish embryos. The discrepancies in identification between the two methods was driven by 2 factors: the lack of interspecific variability in COI for the *Coregonus* genus and limited morphological resolution for *Catostomus* and Cyprinidae.

Acknowledgements

This research was produced in collaboration with Bruce Power. We would also like to thank Mitacs and NSERC for providing funding for this research.

Conflicts of Interest

Dr.Boreham reports other financial contribution from Bruce Power during the conduct of this study.

Author's Contributions and Disclosure

Emily Hulley: Emily was the corresponding author. She was responsible for the majority of the revisions following the original manuscript submission.

Natalie Taylor: Natalie was responsible for conducting research on the larval fish and fish eggs. She was responsible for DNA barcoding, analyzing the compiled data from the morphological identifications and DNA barcoding. She also wrote the original manuscript.

Andrew Zarnke: Andrew contributed to sample processing and manuscript revisions.

Christopher Somers: Chris was a collaborator and was responsible for training Natalie in the techniques used. He was also an editor for the manuscript.

Richard Manzon: Richard was a collaborator involved in providing scientific advice on the project. He was also an editor for the manuscript.

Joanna Wilson: Joanna was a collaborator and she provided scientific advice for the project and was also an editor for the manuscript.

Douglas Boreham: Doug was the academic supervisor for Natalie. He provided advice in terms of data collection and analysis for DNA barcoding. He assisted in the editing of the manuscript.

All authors have approved the final article for submission.

Role of the Funding Source

Funding for this study was provided by Bruce Power, Mitacs and a large NSERC CRD team grant (award # 433617-12). Bruce Power was involved in the collection of specimens and was provided their approval for this article prior to its submission for publication. Mitacs and NSERC had no role in the preparation of this article.

References

April, J., Mayden, R.L., Hanner, R.H., and Bernatchez, P. 2011. Genetic calibration of species diversity among North America's freshwater fishes. Proceedings of the National Academy of Science 108, 10602-10607.

Ardura, A., Linde, A. R., Moreira, J. C., and Garcia-Vazquez, E. 2010. DNA barcoding for conservation and management of Amazonian commercial fish. Biol. Cons. 143(6), 1438-1443.

Austin, M., Buffett, D., Nicolson, D., Stevens, V. and Scudder, G. G. E. 2008. Takin nature's pulse: the status of biodiversity in British Columbia. Biodiversity B.C. Biodiversity B.C. Victoria, BC. 268 p.

Barbuto, M., Galimberti, A., Ferri, E., Labra, M., Malandra, R., Galli, P., and Casiraghi, M. 2010. DNA barcoding reveals fraudulent substitutions in shark seafood products: the Italian case of "palombo" (Mustelus spp.). Food Res. Int. 43(1), 376-381.

Blood, D.M., and Matarese, A.C. 2010. Larval development and identification of the genus triglops. NOAA (National Oceanic and Atmospheric Administration) Professional Paper NMFS, 1-50.

Burkhead, N. M. 2012. Extinction rates in North American freshwater fishes, 1900-2010. BioScience. 62(9), 798-808.

Bruce Power. 2005. Bruce A refurbishment for life extension and continued operations project. Environmental assessment study report volume 1: Main report.

COSEWIC 2015. Committee on the status of endangered wildlife in Canada. 2002 Edition. (Government of Canada).

Fuiman, L.A., and Witman, D.C. 1979. Descriptions and Comparisons of Catostomid Fish Larvae: Catostomus catostomusandMoxostoma erythrurum. Trans. Am. Fish Soc. 108, 604-619.

Fuiman, L.A., Conner, J.V., Lathrop, B.F., Buynak, G.L., Snyder, D.E., and Loos, J.J. 1983. State of the Art of Identification for Cyprinid Fish Larvae from Eastern North America. Trans. Am. Fish Soc. 112, 319-332.

Galimberti, A., Mattia, F. D., Losa, A., Bruni, L., Federici, I., Casiraghi, M., Martellos, S., and Labra, M. 2013. DNA barcoding as a new tool for food traceability. Food. Res. Int. 50(1), 55-63.

Gouy, M., Guindon, S., and Gascuel, O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221-224.

Graham, C.F., Eberts, R.L., Morgan, T.D., Boreham, D.R., Lance, S.L., Manzon, R.G., Mertino, J.A., Rogers, S.M., Wilson, J.Y., and Somers, C.M. 2016. Fine-scale ecological and genetic population structure of two whitefish (coregoninae) species in the vicinity of industrial thermal emissions.PDF. PLoS One 11, e0146656.

Hajibabae, M., Singer, G. AC. Clare, E. L., and Herbert P. DN. 2007. Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. BMC Biol. 5, 24.

Hebert, P.D., and Gregory, T.R. 2005. The promise of DNA barcoding for taxonomy. Syst. Biol. 54, 852-859.

Hubert, N., Hanner, R., Holm, E., Mandrak, N.E., Taylor, E., Burridge, M., Watkinson, D., Dumont, P., Curry, A., Bentzen, P., et al. 2008. Identifying Canadian freshwater fishes through DNA barcodes. PLoS One 3, e2490.

Ivanova, N.V., Zemlak, T.S., Hanner, R.H., and Hebert, P.D.N. 2007. Universal primer cocktails for fish DNA barcoding. Mol. Ecol. Notes 7, 544-548.

Kay, L.K., Wallus, R., and Yeager, B.L. 1994. Reproductive biology and Early life history of fishes in the Ohio River Drainiage, Volume 2: Catostomidae, (Chattanooga, Tennessee: Tennessee Valley Authority).

Khan, N.Y., and Faber, D.J. 1974. A comparison of the larvae of the deepwater and fourhorn sculpin, Myoxocephalus quadricornis L. from North America. I. Morphological development. The Early Life History of Fish, 703-712.

Ko, H.L., Wang, Y.T., Chiu, T.S., Lee, M.A., Leu, M.Y., Chang, K.Z., Chen, W.Y., and Shao, K.T. 2013. Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. PLoS One 8, e53451.

Kochzius, M., Seidel, C., Antoniou, A., Botla, S.K., Campo, D., Cariani, A., Vazquez, E.G., Hauschild, J., Hervet, C., Hjorleifsdottir, S., et al. 2010. Identifying Fishes through DNA Barcodes and Microarrays. PLoS One 5, e12620.

Krantzberg, G. 2006. A validation of ecological services in the Great Lakes basin ecosystem to sustain healthy communities and a dynamic economy. Ontario Ministry of Natural Resources. 8 p.

Meyer, C.P., and Paulay, G. 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 3, e422.

Moura, T., Silvia, M., Figueiredo, I., Neves, A., Muñoz, P., Coelho, M., and Gordo, L. 2008. Molecular barcoding of north-east Atlantic deep-water sharks: species identification and applicaiton to fisheries managment and conservation. Mar. Freshw. Res. 59, 214-223.

Nolte, A. W., Renuat, S., and Bernatchez, L. 2009. Divergence in gene regulation at young life history stages of whitefish (Coregonus sp.) and the emergence of genomis isolation. BMC Evol. Biol. 9, 59.

O'Neill, J.E. 2005. Environmental assessment study report. In Bruce A refurbishment for life extension and continued operations project environmental assessment, Volume 1, T. Brown, ed. (Bruce Power), pp. 5-51 - 55-57.

Ontario. 2017. Species at risk in Ontario. Date accessed: July 21, 2017 from https://www.ontario.ca/environment-and-energy/species-risk-ontario-list

Overdyk, L.M., Braid, H.E., Crawford, S.S., and Hanner, R.H. 2015. Extending DNA barcoding coverage for Lake Whitefish (Coregonus clupeaformis) across the three major basins of Lake Huron. De Gruyter open 3, 59-65.

Pinto, A. D., Pinto, P. D., Terio, V., Bozoo, G., Bonerba, E., Ceci, E., and Tantillo, G. 2013. DNA barcoding for detecting market substitution in salted cod fillets and battered cod chunks. Food Chem. 141(3), 1757-1762.

Renaut, S., Nolte, A. W., and Bernatchez, L. 2009. Gene expression divergence and hybrid misexpression between lake whitefish species pairs (Coregonus spp. Salmonidae). Mol. Biol. Evol. 26(4), 925-936.

Richardson, D.E., Vanwye, J.D., Exum, A.M., Cowen, R.K., and Crawford, D.L. 2007. High-throughput species identification: from DNA isolation to bioinformatics. Mol. Ecol. Notes 7, 199-207.

Richardson, S.L., and Washington, B.B. 1980. Guide to the identification of some sculpin (cottidae) larvae from marine and brackish waters off oregon and adjacent areas in the northease pacific. NOAA (Natl Oceanic Atmos Adm) Tech. Rep. NMFS. (Natl Mar Fish Serv) Circ. 430, 1-61.

Sajdak, S.L., and Phillips, R.B. 1995. Phylogenetic relationships among Coregonus species inferred from teh DNA seqeunce of the first internal transcribed spacer (ITS1) of ribosomal DNA. Can. J. Fish. Aquat. Sci. 54, 1494-1503.

Schlei, O.L., Crete-Lafreniere, A., Whiteley, A.R., Brown, R.J., Olsen, J.B., Bernatchez, L., and Wenburg, J.K. 2008. DNA barcoding of eight North American coregonine species. Mol. Ecol. Resour. 8, 1212-1218.

Smith C. 2016 Environmental Monitoring Program Report. Bruce Power, 2017.

Snyder, D.E. 1998. Catostomid larvae and early juveniles of the Rio Grande basin in Colorado. (Colorado Division of Wildlife), pp. 1-44.

Strauss, R.E., and Bond, C.E. 1990. Taxonomic Methods: Morphology. In Methods for Fish Biology, C. Schreck and P. Moyle, eds. (Maryland: American Fisheries Society), pp. 109-140.

Taylor, H. R., and Harris, W. E. 2012. An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. Mol. Ecol. Resour. 12, 377–388.

Teletchea, F. 2009. Molecular identification methods of fish species- reassessment and possible apppications. Rev. Fish Biol. Fish. 19, 265-293.

Thome, C., Mitz, C., Somers, C.M., Manzon, R.G., Boreham, D.R., and Wilson, J.Y. (2016). Incubation of Lake Whitegish (Coregonus clupeaformis) embryos in cooling water discharge and the impacts of fluctuating thermal regimes on development. Can. J. Fish Aquat. Sci. 0, 1-35.

Valdez-Moreno, M., Quinta-Lizama, C., Gomez-Lozano, R., and Garcia-Rivas, M. C. 2012. Monitoring an alien invasion: DNA barcoding and the identification of lionfish and their prey on coral reefs of the Mexican Caribbean. PLoS One. 7(6), e36636.

Valdez-Moreno, M., Vásquez-Yeomans, L., Elías-Gutiérrez, M., Ivanova, N.V., and Hebert, P.D.N. 2010. Using DNA barcodes to connect adults and early life stages of marine fishes from the Yucatan Peninsula, Mexico: potential in fisheries management. Mar. Freshwater Res. 61, 655-671.

Ward, R. D., Holmes, B. H., White, W. T., and Last, P. R. 2008. DNA barcoding Australasian chondrichthyans: results and potential uses in conservation. Mar. Freshwater Res. 59(1), 57-71.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., and Hebert, P.D. 2005. DNA barcoding Australia's fish species. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360, 1847-1857.

Wong, E. H-K., and Hanner, R. H. 2008. DNA barcoding detects market substitution in North American seafood. Food Res. Int. 41(1), 828-837.

Appendix B: Supplementary Tables

Supplementary Table 1. List of sample source, tissue-type, and number of individuals per group.

Species	# of Samples	Sample Type	Received From
Smallmouth Bass	10	fin clip	Dr. Robert Humston, Washington and Lee University
	5	Muscle	Dr. Louis Bernatchez, Universite Laval
Largemouth Bass	6	Muscle	Dr. Nancy Denslow, University of Florida
Spottail Shiner	13	Muscle	Dr. James Johnson, USGS
Round Whitefish	6	Muscle	Dr. Joanna Wilson, McMaster University
	5	Liver	Dr. James Johnson, USGS
Pygmy Whitefish	3	Muscle	Dr. Hernan Lopez-Fernandez, Royal Ontario Museum
Brook Trout	4	Muscle	Dr. Louis Bernatchez, Universite Laval
Brook Trout CONs	8	Muscle	Dr. Eric B. (Rick) Taylor, University of British Columbia
Lake Whitefish	5	Muscle	Dr. Joanna Wilson, McMaster University
Lake Whitefish CONs	5	Muscle	Dr. Hernan Lopez-Fernandez, Royal Ontario Museum
Deepwater Sculpin	5	Muscle	Dr. Hernan Lopez-Fernandez, Royal Ontario Museum
Fourhorn Sculpin	2	Muscle	Dr. Hernan Lopez-Fernandez, Royal Ontario Museum
Rainbow Smelt	14	Muscle	Dr. Doug Watkinson, Department of Fisheries and Oceans
Yellow Perch	11	muscle	Andrew Zarnke, Laurentian University
Yellow Perch CONs	7	muscle	Dr. Hernan Lopez-Fernandez, Royal Ontario Museum

Supplementary Table 2. Sequent alignment comparison of the primer/probe sets with the corresponding species of interest, control species (same genus) and non-target species (different genus) to determine specificity**.** The number of base pair mismatches are highlighted in red and also tabulated.

Supplementary Table 2.1. Smallmouth Bass (*Micopterus dolomieu*) primer-probe set

	Forward Primer		**Reverse Compl. Primer**		**Internal Oligo**	
	TCTTTCCTTCTCCTGCTCGC		**TTCATCTTGCGGGTGTCTCC**		**GCTGGAGCTGGCACTGGGTG**	
	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches
Micropterus dolomieu	TCTTTCCTTCTCCTGCTCGC	0	TTCATCTTGCGGGTGTCTCC	0	GCTGGAGCTGGCACTGGGTG	0
Coregonus clupeaformis	TCCTTTCTCCTTCTCCTGGC	6	TCCACTTAGCTGGTATTTCC	7	GCCGGTGCCGGCACAGGATG	5
Myoxocephalus thompsonii	TCCTTTCTACTTCTCTTAGC	7	TACATCTAGCAGGAATCTCT	6	GCAGGGGCAGGAACCGGGTG	5
Notropis hudsonius	TCATTTCTACTATTATTAGC	8	TTCACCTAGCAGGTGTCTCA	4	GCTGGGGCTGGAACAGGATG	4
Osmerus mordax	TCCTTTCTTCTTCTCTTAGC	6	TTCACCTTGCGGGGATCTCC	2	GCAGGCGCCGGGACTGGTTG	5
Perca flavescens	TCTTTCCTTCTCCTCCTTGC	2	TACACCTTGCGGGGATCTCC	4	GCCGGAGCTGGTACCGGATG	4
Prosopium cylindraceum	TCCTTTCTTCTTCTCCTGGC	5	TACACTTAGCTGGTATTTCC	7	GCCGGCGCCGGCACAGGATG	5
Salvelinus fontinalis	TCCTTTCTACTTCTCCTGGC	6	TACATTTAGCTGGCATTTCC	7	GCCGGCGCCGGTACGGGGTG	5
Largemouth Bass (*M. salmoides*)	TCCTTCCTTCTCCTGCTCGC	1	TTCACCTTGCTGGTGTCTCC	2	GCCGGGGCTGGCACTGGGTG	2

Supplementary Table 2.2 Spottail shiner (*Notropis hudsonius*) primer/probe set

	Forward Primer		Reverse Compl. Primer		Internal Oligo	
	CTATTATTAGCTTCTTCTGGGGTTG		GGGCGCATCAGTAGACCTC		GCAGGCAATCTTGCCCACGC	
	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches
Notropis hudsonius	CTATTATTAGCTTCTTCTGGGGTTG	0	GGGCGCATCAGTAGACCTC	0	GCAGGCAATCTTGCCCACGC	0
Coregonus clupeaformis	CTTCTCCTGGCCTCGTCCGGAGTTG	9	AGGAGCCTCCGTCGATTTA	8	GCAGGCAACCTCGCCCACGC	2
Micropterus dolomieu	CTCCTGCTCGCCTCTTCCGGGGTCG	8	AGGAGCATCCGTTGACCTA	5	GCCGGCAACCTGGCCCATGC	4
Myoxocephalus thompsonii	CTTCTCTTAGCCTCTTCGGGGGTTG	5	GGGAGCCTCTGTTGACCTA	5	GCCGGAAACCTGGCCCACGC	4
Osmerus mordax	CTTCTCTTAGCTTCTTCCGGGGTTG	4	GGGAGCTTCCGTAGATTTA	6	GCTGGCAATTTGGCCCACGC	3
Perca flavescens	CTCCTCCTTGCTTCCTCAGGAGTTG	8	TGGAGCATCTGTTGATTTA	7	GCTGGGAACTTAGCACATGC	7
Prosopium cylindraceum	CTTCTCCTGGCCTCATCCGGAGTTG	9	AGGGGCCTCCGTTGACTTA	6	GCAGGCAACCTCGCTCACGC	3
Salvelinus fontinalis	CTTCTCCTGGCTTCGTCCGGAGTTG	8	AGGAGCTTCCGTTGATTTA	8	GCTGGGAACCTCGCCCACGC	4
Pugnose Shiner (*N. anogenus*)	CTATTATTAGCCTCTTCTGGCGTTG	2	AGGAGCATCAGTAGACCTC	2	GCGGGTAATCTTGCCCATGC	3
Emerald Shiner (*N. atherinoides*)	TTACTATTAGCCTCCTCTGGTGTTG	5	AGGAGCGTCAGTAGACCTA	4	TCAGGAAACCTTGCCCACGC	3
Bridle Shiner (*N. bifrenatus*)	TTACTATTAGCCTCTTCTGGAGTTG	4	AGGAGCATCAGTAGATCTT	4	GCAGGTAACCTTGCCCATGC	3
River Shiner (*N. blennius*)	TTATTACTAGCCTCTTCTGGAGTTG	4	AGGCGCATCAGTAGACCTT	2	GCAGGCAACCTTGCCCACGC	1
Ghost Shiner (*N. buchanani*)	CTACTACTAGCCTCTTCTGGTGTTG	4	AGGAGCGTCAGTAGACCTA	4	GCGGGTAACCTTGCCCACGC	3
Bigmouth Shiner (*N. dorsalis*)	TTATTACTAGCTTCTTCAGGTGTCG	5	AGGAGCGTCAGTCGACCTC	4	GCAGGCAATCTCGCTCACGC	2
Blackchin Shiner (*N. heterodon*)	TTACTCCTAGCCTCTTCTGGTGTTG	6	AGGAGCATCAGTAGACCTT	3	GCAGGTAATCTTGCCCATGC	2
Blacknose Shiner (*N. heterolepis*)	TTACTATTAGCCTCTTCTGGTGTTG	4	AGGAGCATCAGTAGACCTC	2	GCAGGTAACCTTGCCCACGC	2
Carmine Shiner (*N. percobromus*)	CTACTACTAGCCTCTTCCGGTGTTG	5	AGGAGCATCAGTAGACCTA	3	TCAGGGAACCTTGCCCACGC	3
Silver Shiner (*N. photogenis*)	TTACTATTAGCCTCTTCTGGCGTTG	4	AGGAGCATCAGTAGACCTC	2	GCGGGTAATCTTGCCCACGC	2
Rosyface Shiner (*N. rubellus*)	TTATTACTAGCCTCTTCTGGTGTTG	4	AGGAGCATCAGTAGACCTA	3	TCAGGAAACCTTGCCCATGC	4
Sand Shiner (*N. stramineus*)	CTACTGCTAGCCTCTTCTGGTGTCG	6	AGGGGCATCTGTAGATCTC	4	GCAGGCAATCTTGCCCACGC	0
Weed Shiner (*N. texanus*)	CTACTGCTAGCCTCTTCTGGTGTTG	5	AGGGGCATCAGTAGACCTT	3	GCAGGTAACCTCGCTCACGC	4
Mimic Shiner (*N. volucellus*)	CTACTACTAGCCTCTTCTGGTGTTG	4	AGGAGCGTCAGTAGACCTA	4	GCGGGTAACCTTGCCCACGC	3

Supplementary Table 2.3 Round Whitefish (*Prosopium cylindraceum*) primer/probe set

	Forward Primer		**Reverse Compl.Primer**		**Internal Oligo**	
	AATGTAATCGTCACGGCCCA		**CCCGATATAGCATTCCCCCG**		**TGACTAATTCCCCTTATGATCGGAGCA**	
	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches
Prosopium cylindraceum	AATGTAATCGTCACGGCCCA	0	CCCGATATAGCATTCCCCCG	0	TGACTAATTCCCCTTATGATCGGAGCA	0
Coregonus clupeaformis	AATGTGATCGTCACGGCCCA	1	CCCGACATGGCATTTCCCCG	3	TGATTAATCCCACTTATAATCGGGGCC	6
Micropterus dolomieu	AATGTAATTGTTACAGCGCA	4	CCCGACATAGCATTCCCTCG	2	TGACTTATCCCCCTAATGATCGGTGCC	5
Myoxocephalus thompsonii	AACGTAATTGTTACAGCTCA	5	CCTGACATGGCTTTTCCCCG	5	TGACTCATCCCCTTAATGATTGGGGCC	6
Notropis hudsonius	AACGTTATCGTTACTGCCCA	4	CCTGATATAGCATTTCCACG	3	TGACTTGTACCTCTAATGATCGGAGCA	5
Osmerus mordax	AATGTTATCGTCACCGCGCA	3	CCAGATATGGCCTTCCCTCG	4	TGGCTCATCCCCCTTATGATTGGGGCC	6
Perca flavescens	AACGTAATTGTTACAGCACA	5	CCTGACATAGCTTTCCCTCG	4	TGACTAATTCCACTTATGATCGGTGCC	3
Salvelinus fontinalis	AACGTAATCGTAACAGCCCA	3	CCAGACATAGCATTCCCTCG	3	TGATTAATTCCTCTAATAATTGGAGCC	6
Pygmy Whitefish (*P. coulterii*)	AATGTGATCGTTACAGCCCA	3	CCCGATATAGCATTTCCCCG	1	TGATTAATCCCCCTTATGATTGGGGCA	4
Mountain Whitefish (*P. williamsoni*)	AATGTAATCGTTACAGCCCA	2	CCCGATATAGCATTTCCCCG	1	TGACTAATCCCCCTTATGATCGGAGCA	1

Supplementary Table 2.4 Brook Trout (*Salvelinus fontinalis*) primer/probe set

	Forward Primer		**Reverse Compl.Primer**		**Internal Oligo**	
	CGGTACGGGGTGAACAGTTT		**CCCTACATTTAGCTGGCATTTCC**		**CTCGCCCACGCAGGAGCTTC**	
	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches
Salvelinus fontinalis	CGGTACGGGGTGAACAGTTT	0	CCCTACATTTAGCTGGCATTTCC	0	CTCGCCCACGCAGGAGCTTC	0
Coregonus clupeaformis	CGGCACAGGATGAACAGTCT	4	CCCTCCACTTAGCTGGTATTTCC	3	CTCGCCCACGCAGGAGCCTC	1
Micropterus dolomieu	TGGCACTGGGTGAACTGTCT	5	CTCTTCATCTTGCGGGTGTCTCC	8	CTGGCCCATGCAGGAGCATC	3
Myoxocephalus thompsonii	AGGAACCGGGTGGACAGTAT	5	CCTTACATCTAGCAGGAATCTCT	6	CTGGCCCACGCGGGAGCCTC	3
Notropis hudsonius	TGGAACAGGATGAACTGTCT	6	CTCTTCACCTAGCAGGTGTCTCA	9	CTTGCCCACGCGGGCGCATC	4
Osmerus mordax	CGGGACTGGTTGAACAGTCT	4	CTCTTCACCTTGCGGGGATCTCC	8	TTGGCCCACGCGGGAGCTTC	3
Perca flavescens	TGGTACCGGATGAACTGTTT	4	CTTTACACTTAGCAGGGGTTTCC	6	TTAGCACATGCTGGAGCATC	6
Prosopium cylindraceum	CGGCACAGGATGAACAGTGT	4	CCCTACACTTAGCTGGTATTTCC	2	CTCGCTCACGCAGGGGCCTC	3
Arctic Char (*S. alpinus*)	CGGTACGGGATGGACAGTCT	3	CCCTTCATTTAGCTGGCATTTCC	1	CTCGCCCACGCAGGGGCCTC	2
Bull Trout (*S.confluentus*)	CGGTACGGGATGGACAGTCT	3	CCCTTCATTTAGCTGGCATTTCC	1	CTCGCCCACGCAGGGGCCTC	2
Dolly Varden (*S. malma*)	CGGTACGGGATGGACAGTCT	3	CCCTTCATTTAGCTGGCATTTCC	1	CTCGCCCACGCAGGAGCCTC	1
Lake Trout (*S. namaycush*)	CGGTACGGGATGAACAGTCT	2	CTCTTCATTTAGCTGGCATTTCC	2	CTCGCCCACGCAGGGGCCTC	2

Supplementary Table 2.5 Lake Whitefish (*Coregonus clupeaformis*) primer/probe set

	Forward Primer		Reverse Compl.Primer		Internal Oligo	
	TCTCCCTCCACTTAGCTGGT		ACCCCTCTTTTTGTCTGGGC		TTCCTCTATCTTGGGGGCCGTT	
	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches
Coregonus clupeaformis	TCTCCCTCCACTTAGCTGGT	0	ACCCCTCTTTTTGTCTGGGC	0	TTCCTCTATCTTGGGGGCCGTT	0
Micropterus dolomieu	TCTCTCTTCATCTTGCGGGT	6	ACACCCCTGTTTGTTTGGTC	5	CTCCTCCATCCTAGGGGCCATC	6
Myoxocephalus thompsonii	TCTCCTTACATCTAGCAGGA	5	ACCCCTCTATTCGTGTGATC	5	CTCTTCGATCCTCGGAGCAATC	9
Notropis hudsonius	TCTCTCTTCACCTAGCAGGT	4	ACACCTCTTTTCGTATGAGC	4	CTCATCAATTCTAGGGGCAGTT	7
Osmerus mordax	TCTCTCTTCACCTTGCGGGG	6	ACCCCCTTATTTGTCTGAGC	4	CTCCTCTATTCTCGGGGCAATT	6
Perca flavescens	TCTCTTTACACTTAGCAGGG	5	ACTCCCTTGTTCGTATGGGC	5	TTCCTCAATTCTAGGTGCTATT	7
Prosopium cylindraceum	TCTCCCTACACTTAGCTGGT	1	ACACCCCTTTTTGTTTGAGC	4	TTCCTCTATTTTAGGAGCCGTT	3
Salvelinus fontinalis	TTTCCCTACATTTAGCTGGC	4	ACCCCACTTTTTGTTTGAGC	3	TTCCTCAATTTTAGGAGCCATT	5
Cisco (*C. artedi*)	TCTCCCTCCACTTAGCTGGT	0	ACCCCTCTTTTTGTCTGAGC	1	TTCCTCTATCTTAGGAGCCGTT	2
Arctic Cisco (*C. autumnalis*)	TCTCCCTCCACTTAGCTGGT	0	ACCCCTCTTTTTGTCTGAGC	1	TTCCTCTATCTTAGGAGCCGTT	2
Bloater (*C. hoyi*)	TCTCCCTCCACTTAGCTGGT	0	ACCCCCCTGTTTGTCTGAGC	3	TTCCTCTATCTTAGGAGCCGTT	2
Atlantic Whitefish (*C. huntsmani*)	TCTCCCTCCACCTAGCTGGT	1	ACCCCTCTTTTTGTTTGAGC	2	TTCCTCTATCTTGGGAGCCGTT	1
Kiyi (*C. kiyi*)	TCTCCCTCCACTTAGCTGGT	0	ACCCCTCTTTTTGTCTGAG	1	TTCCTCTATCTTAGGAGCCGTT	2
Bering Cisco (*C. laurettae*)	TCTCCCTCCACTTAGCTGGT	0	ACCCCTCTTTTTGTCTGAGC	1	TTCCTCTATCTTAGGAGCCGTT	2
Broad Whitefish (*C. nasus*)	TCTCCCTCCACTTAGCTGGT	0	ACCCCTCTTTTTGTCTGGGC	0	TTCCTCTATCTTAGGGGCCGTT	1
Blackfin Cisco (*C. nigripinnis*)	TCTCCCTCCACTTAGCTGGT	0	ACCCCTCTTTTTGTCTGAGC	1	TTCCTCTATCTTAGGAGCCGTT	2
Sardine Cisco (*C. sardinella*)	TCTCCCTCCACTTAGCTGGT	0	ACCCCTCTTTTTGTCTGAGC	1	TTCCTCTATCTTAGGGGCCGTT	1
Shortjaw Cisco (*C. zenithicus*)	TCTCCCTCCACTTAGCTGGT	0	ACCCCTCTTTTTGTCTGAGC	1	TTCCTCTATCTTAGGAGCCGTT	2

Supplementary Table 2.6 Deepwater Sculpin (*Myoxocephalus thompsonii*) primer/probe set

	Forward Primer		**Reverse Compl. Primer**		**Internal Oligo**	
	CTTAGCCTCTTCGGGGGTTG		**TCTCTTCGATCCTCGGAGCA**		**CCACGCGGGAGCCTCTGTTG**	
	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches
Myoxocephalus thompsonii	CTTAGCCTCTTCGGGGGTTG	0	TCTCTTCGATCCTCGGAGCA	0	CCACGCGGGAGCCTCTGTTG	0
Coregonus clupeaformis	CCTGGCCTCGTCCGGAGTTG	5	TTTCCTCTATCTTGGGGGCC	7	CCACGCAGGAGCCTCCGTCG	3
Micropterus dolomieu	GCTCGCCTCTTCCGGGGTCG	5	TCTCCTCCATCCTAGGGGCC	5	CCATGCAGGAGCATCCGTTG	4
Notropis hudsonius	ATTAGCTTCTTCTGGGGTTG	3	TCTCATCAATTCTAGGGGCA	5	CCACGCGGGCGCATCAGTAG	4
Osmerus mordax	CTTAGCTTCTTCCGGGGTTG	2	TCTCCTCTATTCTCGGGGCA	4	CCACGCGGGAGCTTCCGTAG	3
Perca flavescens	CCTTGCTTCCTCAGGAGTTG	6	TTTCCTCAATTCTAGGTGCT	7	ACATGCTGGAGCATCTGTTG	4
Prosopium cylindraceum	CCTGGCCTCATCCGGAGTTG	5	TTTCCTCTATTTTAGGAGCC	7	TCACGCAGGGGCCTCCGTTG	4
Salvelinus fontinalis	CCTGGCTTCGTCCGGAGTTG	6	TTTCCTCAATTTTAGGAGCC	7	CCACGCAGGAGCTTCCGTTG	3
Fourhorn Sculpin (*M. quadricornis*)	CTTAGCCTCTTCGGGGGTTG	0	TCTCTTCAATCCTCGGAGCA	1	CCACGCGGGAGCCTCTGTTG	0

Supplementary Tale 2.7 Rainbow Smelt (*Osmerus mordax*) primer/probe set

	Forward Primer		**Reverse Compl. Primer**		**Internal Oligo**	
	CGATTATGATCGGCGGGTTTG		**GATATGGCCTTCCCTCGCAT**		**CCCCCTTATGATTGGGGCCCCA**	
	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches
Osmerus mordax	CGATTATGATCGGCGGGTTTG	0	GATATGGCCTTCCCTCGCAT	0	CCCCCTTATGATTGGGGCCCCA	0
Coregonus clupeaformis	CAATTATGATTGGAGGCTTTG	4	GACATGGCATTTCCCCGAAT	5	CCCACTTATAATCGGGGCCCCC	4
Micropterus dolomieu	CCATTATAATTGGAGGCTTTG	5	GACATAGCATTCCCTCGAAT	4	CCCCCTAATGATCGGTGCCCCC	4
Myoxocephalus thompsonii	CAATCATAATTGGGGGTTTCG	8	GACATGGCTTTTCCCCGAAT	5	CCCCTTAATGATTGGGGCCCCT	3
Notropis hudsonius	CAATTCTTATTGGCGGATTTG	5	GATATAGCATTTCCACGAAT	5	ACCTCTAATGATCGGAGCACCT	7
Perca flavescens	CAATTATGATTGGGGGCTTTG	4	GACATAGCTTTCCCTCGAAT	4	TCCACTTATGATCGGTGCCCCT	5
Prosopium cylindraceum	CAATTATGATTGGAGGATTTG	4	GATATAGCATTCCCCCGAAT	4	TCCCCTTATGATCGGAGCACCC	5
Salvelinus fontinalis	CAATTATGATTGGAGGATTTG	4	GACATAGCATTCCCTCGAAT	4	TCCTCTAATAATTGGAGCCCCA	5
Pacific Rainbow Smelt (*O. dentex*)	CAATCATGATTGGAGGTTTCG	6	GACATGGCCTTCCCCCGTAT	3	CCCCCTTATGATTGGGGCCCCA	0

Supplementary Table 2.8 Yellow Perch (*Perca flavescens*) primer/probe set

	Forward Primer		**Reverse Compl.Primer**		**Internal Oligo**	
	GATCGGTGCCCCTGACATAG		**TTATCCCCCTCTTGCTGGGA**		**AAGCCGGAGCTGGTACCGGA**	
	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches	Aligned Primer	# of mismatches
Perca flavescens	GATCGGTGCCCCTGACATAG	0	TTATCCCCCTCTTGCTGGG	0	AAGCCGGAGCTGGTACCGGA	0
Coregonus clupeaformis	AATCGGGGCCCCCGACATGG	4	CTACCCCCCTCTGGCAGGC	5	AAGCCGGTGCCGGCACAGGA	4
Micropterus dolomieu	GATCGGTGCCCCCGACATAG	1	CTACCCCCCTCTTGCCGGC	4	AAGCTGGAGCTGGCACTGGG	4
Myoxocephalus thompsonii	GATTGGGGCCCCTGACATGG	3	ATACCCTCCCCTTGCCGGA	6	AAGCAGGGGCAGGAACCGGG	5
Notropis hudsonius	GATCGGAGCACCTGATATAG	3	CTACCCCCCACTTGCAGGC	5	AAGCTGGGGCTGGAACAGGA	4
Osmerus mordax	GATTGGGGCCCCAGATATGG	5	CTATCCCCCACTTGCTGGC	3	AAGCAGGCGCCGGGACTGGT	6
Prosopium cylindraceum	GATCGGAGCACCCGATATAG	4	GTATCCCCCACTAGCAGGC	5	AAGCCGGCGCCGGCACAGGA	4
Salvelinus fontinalis	AATTGGAGCCCCAGACATAG	4	TTACCCCCCTCTAGCTGGG	2	AAGCCGGCGCCGGTACGGGG	4
Logperch (*Percina caprodes*)	GATCGGCGCCCCCGATATGG	4	CTACCCGCCTTTAGCGGGA	7	AAGCAGGGGCTGGAACTGGG	5
Channel Darter (*Percina copelandi*)	GATCGGCGCCCCCGACATGG	3	ATACCCACCTCTGGCTGGG	4	AGGCTGGAGCTGGAACCGGA	3
Blackside Darter (*Percina maculata*)	GATTGGTGCCCCCGACATGG	3	CTACCCGCCCCTGGCTGGA	6	AAGCTGGGGCTGGAACCGGA	3
River Darter (*Percina shumardi*)	GATCGGTGCCCCCGACATGG	2	TTACCCGCCTCTGGCCGGA	5	AAGCTGGAGCTGGAACTGGA	3

Supplementary Table 3. qPCR Ct values for single-plex and multiplex analysis with varying DNA amounts (1 pg to 300 µg) per reaction. Each species of interest was assayed using single-plex and multiplex conditions, while the control species from the same genus were assayed using multiplex conditions only. Probe set one consisted of primer/probe sets for following species of interest: smallmouth bass, spottail shiner, round whitefish, and brook trout (Supplementary 3.1). Probe set two consisted of primer/probe sets for following species of interest: lake whitefish, deepwater sculpin, rainbow smelt and yellow perch (Supplementary 3.2). Species that were undetectable were marked as "-".

Supplementary 3.1

Smallmouth Bass Primer-Probe Set					Spottail Shiner Primer-Probe Set						
	Smallmouth Bass		Largemouth Bass			Spottail Shiner		Bigmouth Shiner	Carmine Shiner	Rosyface Shiner	Weed Shiner
[DNA] ng	Single Plexing	Multiplexing			[DNA] ng	Single Plexing	Multiplexing				
300.0	20.87	20.78	-		300.0	16.64	16.73	38.2	-	-	-
100.0	22.90	23.15	-		100.0	18.03	18.07	-	-	-	-
30.00	25.25	25.71	-		30.00	19.70	19.79	-	-	-	-
10.00	28.03	28.19	-		10.00	21.27	21.26	-	-	-	-
3.00	30.67	30.92	-		3.00	23.00	22.99	-	-	-	-
1.00	33.14	33.08	-		1.00	24.77	24.80	-	-	-	-
0.30	36.73	37.28	-		0.30	26.72	26.89	-	-	-	-
0.10	39.55	39.74	-		0.10	29.27	29.29	-	-	-	-
0.030	-	-	-		0.030	31.31	31.47	-	-	-	-
0.010	-	-	-		0.010	34.12	34.66	-	-	-	-
0.003	-	-	-		0.003	35.85	35.63	-	-	-	-
0.001	-	-	-		0.001	36.85	36.46	-	-	-	-
Round Whitefish Primer-Probe Set					**Brook Trout Primer-Probe Set**						
	Round Whitefish		Pygmy Whitefish			Brook Trout		Lake Trout	Arctic Char	Bull Trout	Dolly Varden
[DNA] ng	Single Plexing	Multiplexing			[DNA] ng	Single Plexing	Multiplexing				
300.0	19.8	19.9	35.80		300.0	18.53	18.7	36.4	36.9	34.7	38.5
100.0	22.6	22.5	35.48		100.0	20.36	20.5	38.3	38.5	36.7	39.7
30.00	25.2	25.2	-		30.00	22.11	22.3	-	39.9	38.0	-
10.00	27.6	27.4	-		10.00	23.58	23.9	39.5	-	39.2	-
3.00	30.2	30.1	-		3.00	25.38	25.7	-	-	39.7	-
1.00	32.9	33.1	-		1.00	27.10	27.3	-	-	-	-
0.30	35.3	35.1	-		0.30	28.98	29.2	-	-	-	-
0.10	37.9	39.2	-		0.10	31.04	31.4	-	-	-	-
0.030	-	-	-		0.030	33.09	33.3	-	-	-	-
0.010	-	-	-		0.010	34.99	35.1	-	-	-	-
0.003	-	-	-		0.003	37.90	36.6	-	-	-	-
0.001	-	-	-		0.001	39.48	38.6	-	-	-	-

Supplementary 3.2

Lake Whitefish Primer-Probe Set							Deepwater Sculpin Primer-Probe Set		
	Lake Whitefish		Cisco	Bloater	Kiyi			Deepwater Sculpin	
[DNA] ng	Single Plexing	Multiplexing					[DNA] ng	Single Plexing	Multiplexing
300.00	15.9	16.0	-	-	-		300.00	16.0	16.6
100.00	17.6	17.9	-	-	-		100.00	18.0	18.6
30.00	19.8	19.7	-	-	-		30.00	20.0	20.7
10.00	22.7	22.0	-	-	-		10.00	22.2	22.5
3.00	23.9	24.1	-	-	-		3.00	24.0	24.7
1.00	25.8	26.0	-	-	-		1.00	26.1	26.9
0.30	28.1	28.3	-	-	-		0.30	28.5	30.5
0.10	30.3	30.6	-	-	-		0.10	30.7	31.4
0.030	31.9	31.4	-	-	-		0.030	32.6	33.2
0.010	33.9	34.1	-	-	-		0.010	34.2	35.0
0.0030	35.8	36.1	-	-	-		0.0030	36.5	37.3
0.0010	37.5	37.5	-	-	-		0.0010	38.1	38.2

Yellow Perch Primer-Probe Set							Rainbow Smelt Primer-Probe Set		
	Yellow Perch		Log Perch	Blackside Darter	River Darter			Rainbow Smelt	
[DNA] ng	Single Plexing	Multiplexing					[DNA] ng	Single Plexing	Multiplexing
300.00	17.7	17.5	37.0	35.5	35.7		300.00	19.4	19.5
100.00	19.3	19.3	-	37.1	-		100.00	20.9	21.2
30.00	21.0	20.9	-	36.4	-		30.00	22.3	22.6
10.00	22.6	22.6	-	-	-		10.00	24.2	24.1
3.00	24.5	24.5	-	-	-		3.00	26.1	26.3
1.00	26.5	26.5	-	-	-		1.00	29.5	29.5
0.30	28.8	28.9	-	-	-		0.30	30.9	30.9
0.10	30.8	30.8	-	-	-		0.10	32.9	33.0
0.030	29.4	29.3	-	-	-		0.030	35.2	35.1
0.010	31.3	31.1	-	-	-		0.010	37.9	37.2
0.0030	34.8	33.2	-	-	-		0.0030	38.0	38.7
0.0010	35.7	34.6	-	-	-		0.0010	-	38.6

Supplementary Table 4. Literature review summary of aquatic environmental DNA (eDNA) studies.

Paper	Sampling Sites	Species	Amount Collected	Extraction Method	Storage	Sequencing
Ficetola *et al*. 2008	3 Ponds (1000 - 10,000m²) in France	American bullfrog (*Rana catesbeiana*)	15mL	Centrifuged: 5500g, 35 mins, 6°C DNA extracted from pellet using DNeasy tissue extraction kit (Qiagen)	Added 1.5mL acetate 3M and 33mL ethanol to sample and stored at -20°C until extraction	Species-specific primers (78bp on Species-specific primers (78bp on *cyt b* gene). Samples run on PCR 3-5 times and visualized by gel electrophoresis. One pond sequenced via 454 pyrosequecing technology
Furlan and Gleeson 2016	3 Waterways in Australia Capital Territory	Redfin perch (*Perca fluviatilis*)	2L	Filtered onto 1.2µm pore size glass fiber filters Extraction via PowerWater DNA kit	Filters rolled and placed into 5mL tube and stored at -20°C before extraction	Species-specific primer-probe assays (92bp on 12S rRNA gene) using real-time PCR
Jerde *et al.* 2011	Large river and canal complex in Chicago area	Silver carp (*Hypophthalmichthys molitrix*) Bighead carp (*H.nobilisi*)	2L	Filtered onto 1.52µm pore size glass fiber filters Extraction via PowerWater DNA kit	Filters placed into 50mL tubes and stores at -20°C	PCR and gel electrophoresis Positive bands extracted from gel and bidirectionally sequenced on DNA analyzer
Deiner and Altermatt 2014	Lake Greifensee in Switzerland (Surface area: 8.5km³)	2 invertebrates: *Daphnia longispina* and *Unio tumidus*	900mL	Filtered 300mL onto 0.2252µm pore size glass fiber filters Extraction using cell lysis, phenol chloroform isoamyl procedure	Water samples stored at -20°C	Species-specific primer-probe (12S and COI, bp 157-175), standard PCR and Sanger sequencing
Machler *et al.* 2014	Rivers and lakes in canton of Zurich in Switzerland	6 invertebrates: *Crangonyx pseudogracilis* *Gammarus pulex* *Asellus aquaticus* *Ancylus fluviatilis* *Baetis bucceratus* *Tinodes waeneri*	1L	Filtered 280-300mL onto 0.7-µm glass fiber filter 4 times Extraction by cell lysis, phenol chloroform isoamyl procedure	Water samples stored at -20°C	Species-specific primers (COI), PCR and gel electrophoresis
Goldberg *et al.* 2013	Portneuf River, Idaho	New Zealand mudsnail (*Potamopyrgus antipodarum*)	4L	Filtered onto mixed cellulose ester membranes with 0.45-µm pore size QIAshredder/DNeasy Blood and Tissue DNA extraction kit	Stored filters in 95% ethanol	Species-specific primers (*Cyt b*) and qPCR
Maruyama *et al.* 2014	Southern basin of Lake Biwa in Shiga	Bluegill sunfish (*Lepomis macrochirus*)	15mL	Centrifuged: 1 hour, 4°C, 10,000g Pellet resuspended in 200mL water and	1.5mL of 3M sodium acetate and 30mL ethanol and	Real-time qPCR with species-specific primer/probes (*cyt b*)

				extracted from solution via DNeasy tissue kit (Qiagen)	stored at -40°C	
Doi *et al.* 2015	12 outdoor tanks (450L)	Common carp (*Cyprinus carpio*)	50 mL	Centrifuged: 10,000g, 1 hour, 4°C eDNA extracted from pellet using DNeasy tissue kit (Qiagen)	1.5mL of 3M sodium acetate and 33mL ethanol and stored at -20°C	qPCR vs ddPCR

Supplementary Table 5. Average Ct values of positive and negative controls for qPCR analyse of yellow perch from eDNA samples. 15ng of DNA was used for all reactions.

Species	Number of Samples	Average Ct Value
Yellow Perch	2	21.91±0.31
Smallmouth Bass	2	undetectable
Lake Whitefish	2	undetectable
Rainbow smelt	2	undetectable

Supplementary Table 6. Species identified in Lake Laurentian at each site from the next generation sequencing results. Next generation sequencing counts represent the total number of sequences from a specific species identified from eDNA samples ≥98% homology match.

Species Name	Percent Homology	Sites					
		A	B	C	D	E	F
anopheles annulipes	99	0	0	0	2	1	0
anopheles neivai	98	13	1	0	0	0	0
anopheles rangeli	98	0	0	0	0	2	0
azteca pittieri	98	0	3	0	0	0	0
baetis rhodani	98	7	5	17	14	1	29
batrachoseps attenuatus	98	12	7	5	3	0	5
batrachospermum turfosum	99	1	0	0	0	15	0
brachionus macracanthus	98	0	28	46	58	1	0
caenis amica	98	1	1	1	1	1300	1
caenis punctata	99	0	1	2	0	2219	2
caenis sp.	98	1	1	0	0	991	0
callosobruchus maculatus	100	13	1	24	13	0	4
chaetogaster diastrophus	98	0	0	3	49	1	0
crucella scotiae	98	26	16	3	11	1	17
enchenopa sp. 'carya'	98	2	2	2	1	0	1
euchlanis dilatata	100	5415	9717	11827	6763	336	8470
homo sapiens	98	15	9	0	0	0	0
macrothrix sp.	98	38	9	30	19	76	64
megalurothrips sp.	98	0	0	0	0	21	0
megaselia halterata	98	0	0	0	0	183	0
mus musculus	98	21	1	0	0	0	0
mycetophila alea	98	10	541	3567	167	2	807
mycetophila alea	98	10	541	3567	167	2	807
notropis heterodon	99	75	3	0	0	0	0
ochoterenella sp.	98	12	1	3	0	0	69
pelophylax nigromaculatus	100	7	9	3	3	0	14
perca flavescens	100	201	38	17	34	0	14
phytomyza tetrasticha	98	0	0	0	0	0	3
rotaria macroceros	98	0	0	0	0	0	26
selenops pensilis	98	408	38	308	545	59	269
skistodiaptomus oregonensis	99	44	0	0	0	0	0
synchaeta lakowitziana	99	8	7	1700	1780	1	958
thoracosphaera heimii	99	68	88	389	625	110	59