INCREASING LIMITS OF ENVIRONMENTAL DNA DETECTION USING THE ENDANGERED ATLANTIC WHITEFISH (*COREGONUS HUNTSMANI*) AS A TEST CASE

by

Samantha Beal

Submitted in partial fulfilment of the requirements
for the degree of Masters of Science

at

Dalhousie University
Halifax, Nova Scotia
March 2024

Dalhousie University is located in Mi'kma'ki, the
ancestral and unceded territory of the Mi'kmaq.
We are all Treaty people.

Dedication

To the Atlantic Whitefish.

Table of Contents

List of Tables

List of Figures

Abstract

Environmental (e)DNA analysis offers a non-invasive alternative to traditional aquatic monitoring, but it can struggle to detect species at very low abundances. Here, I developed a novel eDNA marker targeting the *Sma*I-corII SINE and compared its detection capabilities to the mitochondrial marker ND4 using the endangered Atlantic Whitefish (*Coregonus huntsmani*) as a test case. My results showed that the SINE marker, which is 190x and 82x more abundant than ND4 in gDNA and eDNA samples, respectively, enhanced species detection in eDNA samples compared to the mitochondrial marker. I also investigated the use of SINEs to assess genetic diversity within eDNA samples. My analysis revealed 5 *Sma*I-corII variants within Atlantic Whitefish gDNA, all of which were also detected in eDNA samples. Additionally, I identified two variants unique to Lake Whitefish (*Coregonus clupeaformis*) eDNA samples. This study demonstrates that SINEs serve as sensitive eDNA markers with both intra- and interspecific variation. Given that SINEs and other transposable elements are present in most organisms, these findings have implications for supporting the ongoing management of at-risk aquatic species.

List of Abbreviations and Symbols Used

A_E	Allelic richness
AIS	Aquatic invasive species
ASV	Amplicon sequence variant
bp	Base pair
c/µL	Copies per microlitre
COSEWIC	Committee on the Status of Endangered Wildlife in Canada
CP/Cq	Crossing point/cycle number where fluorescence crosses the threshold in qPCR
CTAB	Cetyltrimethylammonium bromide
CV	Coefficient of variation
DFO	The Department of Fisheries and Oceans Canada
DNA	Deoxyribonucleic acid
eDNA	Environmental DNA
EDTA	Ethylenediaminetetraacetic acid
gDNA	Genomic DNA
GE/µL	Genome equivalents per microlitre
H_0	Null hypothesis
H_A	Alternative hypothesis
H_E	Expected heterozygosity
Indels	Insertions and deletions
ITS	Internal transcribed spacer
LOD	Limit of detection
LOQ	Limit of quantification
m	Metres
MGPL	Marine Gene Probe Lab
min	Minutes
mtDNA	Mitochondrial DNA
mt-eDNA	Mitochondrial eDNA
mL/s	Millilitre per second
mM	Millimolar
MYA	Million years ago
NaCl	Sodium chloride
ND4	NADH dehydrogenase 4 subunit
NGS	Next generation sequencing
nM	Nanometre
Ns	Ambiguous bases
NTC	No template control
nuDNA	Nuclear DNA
nu-eDNA	Nuclear eDNA
PCI	Phenol-chloroform-isoamyl alcohol
PCR	Polymerase chain reaction
PET	Polyethylene terephthalate
PVP	Polyvinylpyrrolidone
qPCR	Quantitative PCR
RNA	Ribonucleic acid

s	Seconds
SINE	Short interspersed nuclear element
SNP	Single nucleotide polymorphism
T_A	Annealing temperature
TEs	Transposable elements
TE	Tris-EDTA
tRNA	Transfer RNA
μ	Mean
μL	Microlitre
μM	Micrometre

Acknowledgements

There are many people who made the completion of this thesis possible. To my supervisors, Dr. Paul Bentzen and Dr. Ian Bradbury, thank you for providing the support and guidance needed to complete this project as well as for providing me opportunities to share my findings with the scientific community and grow as a scientist. I also thank Dr. Daniel Ruzzante and Dr. Julie LaRoche for their valuable contributions as members of my ATC and thesis committees, Dr. Andrew Schofield for serving as the external examiner at my ATC, and Dr. Erin Grey for serving as the external examiner for my defence. Thank you Ian Paterson and Matt Penney for all the time spent with me in the lab, taking the time to train me on new techniques and troubleshoot as needed. To Beth Watson, thank you for helping me keep my head above water during the learning curve of bioinformatics, for all your thoughtful insights related to project design, and your ongoing reassurance and encouragement. I thank Sammy Crowley and Emily Yeung for all their time spent listening to my confusions and elations surrounding this project and for helping me work through encountered issues. To all other members of the Bentzen and Bradbury labs, thank you for all your feedback for many thought-provoking discussions.

I would also like to thank Jeremy Broome and his DFO colleagues as well as Coastal Action for their assistance with sampling the Petite Rivière system and to John Batt for providing access to Atlantic Whitefish within the Aquatron my eDNA sampling.

Lastly, I would like to thank my family for their ongoing support as I completed this thesis and for the numerous visits from the Coffee Fairy which aided in the writing of this thesis. And to Vitek, your constant encouragement and perspective have been invaluable. Thank you for being an essential part of this endeavor.

This work was funded by the Patrick F. Lett Bursary (Dalhousie University), the Faculty of Graduate Studies Departmental Allocation Funding (Dalhousie University), and Discovery Grants from The Natural Sciences and Engineering Research Council of Canada (NSERC).

Chapter 1 – Introduction

1.1 Environmental DNA

Extinction is a natural phenomenon: of the estimated four billion species to have ever existed on Earth, 99% no longer exist (Barnosky et al. 2011). While common, extinction has typically been balanced by speciation. There have been exceptions to this balance, namely five periods of elevated extinction rates – the five mass extinction events. Contemporary global biodiversity is in crisis with current species extinction rates higher than those experienced during the previous mass extinction events (Barnosky et al. 2011). Most current observed species loss can be attributed to anthropogenic causes such as habitat destruction and fragmentation, the introduction of non-native species to new environments, deliberate human destruction of species, and the human-induced climate crisis (Barnosky et al. 2011). The loss of species decreases overall biodiversity, which in turn may affect ecosystems functioning and the goods and services they provide (Cardinale et al. 2012).

Although biodiversity loss is a global concern, freshwater ecosystems and species are particularly vulnerable (Dudgeon et al. 2006). Covering less than 1% of Earth's surface, freshwater ecosystems house nearly one-in-three described vertebrate species, including around 17,800 fish species (Tickner et al. 2020). The high level of endemism present within freshwater ecosystems further contributes to their vulnerability regarding biodiversity loss; for highly endemic species, a local disappearance is a step toward to extinction. Compared to those in marine (Grooten and Almond 2018) or terrestrial ecosystems, populations of freshwater vertebrate species have declined at twice the rate since 1970 (Dudgeon 2010). The causes of such decline have been well documented. Freshwater systems are often near human settlement and are therefore exposed to a multitude of anthropogenic stressors including, but not limited to, pollution, degradation and destruction of habitat, over-extraction of water, over-exploitation of commercially viable species, the impacts of introduced species, and climate change (Dudgeon 2010). These threats may occur concurrently and in an additive manner (Williams-Subiza and Epele 2021); for example, climate change may exacerbate eutrophication due to the higher water temperatures increasing nutrient release from lake sediment, thereby increasing algal growth (Rodgers 2021). Additionally, fresh water is situated within terrestrial ecosystems, and acts as a sink for waste runoff while having limited capacity to dilute these contaminants (Dudgeon 2010). The conservation and management of these vulnerable ecosystems requires accurate assessment of the biodiversity changes occurring within (Zhang et al. 2020).

Traditional detection methods for freshwater fishes, which often involve netting or trapping, are time- and labour-intensive as well as invasive (Sakata et al. 2021). The collection of DNA from environmental samples, a tool known as environmental DNA (eDNA), is a promising alternative to such approaches (Fediajevaite et al. 2021). While there has been great interest in eDNA application, less focus has been dedicated to understanding the mechanics of eDNA generation and decay. As organisms exist within their environment, they contribute DNA to the surrounding area through secretion of mucus or saliva, or cells and tissues flaking away via exfoliation, excretion, and reproductive activities (Barnes and Turner 2016). In addition to eDNA from live animals, decomposition is a source of eDNA as well. Once in the environment, eDNA exists as both intra- and extracellular DNA (Barnes and Turner 2016). It has been proposed eDNA first exists within whole cells shed from multicellular organisms and that those cells then break down, releasing the DNA within (Barnes and Turner 2016). When compared to traditional sampling methods, eDNA sampling can be more cost-effective and more sensitive for detecting aquatic species; however, the magnitude of difference varies among taxa (Fediajevaite et al. 2021). eDNA tools can be particularly beneficial when working in remote or hard-to-access locations which cannot support a traditional fish survey whether due to time, logistical, or equipment-related constraints (Nolan et al. 2023).

Though the roots of eDNA analysis can be found in the older fields of ancient DNA, forensics, and microbial analyses, detection of aquatic eDNA from a multicellular animal was first reported in the late 2000s (Ficetola et al. 2008). Since then, the use of eDNA has rapidly expanded, with the number of eDNA publications growing from four in 2012 to 28 in 2018 and 124 in 2021 (Takahashi et al. 2023). This growth has been accompanied by a diversification of methods and techniques used. Broadly, eDNA studies typically fall into one of two categories. The first, targeted detection of a focal species using taxon-specific primers, has been successfully employed to detect rare (Jerde et al. 2011), endangered (Weltz et al. 2017), and invasive species (Rojahn et al. 2021). This method usually uses quantitative PCR (qPCR) for detection and quantification, where a fluorescent probe binds to the target DNA in the PCR assay. The copy number of target fragments can be quantified by creating a standard curve from DNA standards of known concentrations and comparing the point at which each crosses the background noise threshold to when the unknown sample crosses the threshold. The second category is metabarcoding, in which multiple species are detected in a sample using amplification by "universal" primers followed by DNA sequencing and bioinformatic identification of taxa (Sakata et al. 2021). Metabarcoding has successfully been used to characterize the community composition of both marine (Gold et al. 2021) and freshwater (Xie et al. 2021) habitats.

Though eDNA has the potential to transform how aquatic biodiversity is assessed, it is not without challenges and limitations. While there is much interest in quantifying species abundance from eDNA analysis, these estimates are confounded by the expectation that individuals release eDNA into the environment at different rates (Sigsgaard et al. 2020). Further, how eDNA shedding rate changes during different activities is not well understood (Klymus et al. 2015). Analysis of intraspecific genetic variation using eDNA samples is currently hindered by the inability to distinguish between individuals (Adams et al. 2019). Amplification errors, including allelic drop out, false alleles, and PCR inhibitors can lead to false negatives and are another concern (Adams et al. 2019; Fediajevaite et al. 2021). Further compounding the issue, the timescale over which eDNA remains detectable is heavily influenced by environmental conditions and there remains uncertainty regarding both abiotic and biotic influences on eDNA degradation (Beng and Corlett 2020). As eDNA analyses vary greatly among species and systems, there is little standardization across the literature on optimized practices (Takahashi et al. 2023). Additionally, eDNA methods carry the risk of false positives through amplification of DNA not representative of the contemporary biota in the area, such as DNA resuspended from sediment, transferred from one sampling site to another, or from contaminated equipment (Takahashi et al. 2023).

As eDNA analyses continue to be implemented in management and conservation plans, there is a need to better understand how methodological choices influence obtained results (Alexander et al. 2023). Much focus has been directed toward ensuring the specificity of genetic markers (hereafter called "markers") in targeted studies and the generality of metabarcoding markers; however, less attention has been dedicated to reporting marker sensitivity (Xia et al. 2021), defined as the lowest amount of DNA in a sample that can be detected with 95% probability (Klymus et al. 2020). The detection of low-abundance species via eDNA requires accurate and sensitive assays, particularly when targeting invasive or endangered species, cases where false negative results can have major implications by contributing to the delay or cessation of mitigation or conservation strategies, respectively (Xia et al. 2021). As markers are developed and published, they become available for use anywhere the target species resides. Therefore, it is imperative that marker limitations are well established and reported.

1.2 Atlantic Whitefish (*Coregonus huntsmani*)

Atlantic Whitefish (*Coregonus huntsmani*) is an endangered salmonid endemic to Nova Scotia, Canada (Edge 1984). One of only four freshwater fishes endemic to Canada (Enns et al. 2020), Atlantic Whitefish is also the oldest member of the *Coregonus* lineage with estimates of their divergence from the common *Coregonus* ancestor put at over 14 million years ago (Crête-Lafrenière et al. 2012). Though an ancient species, Atlantic Whitefish was only recently described (Scott 1987), having previously been misidentified as another whitefish species known to reside across Nova Scotia, Lake Whitefish (*Coregonus clupeaformis*).

Historically, the species was documented in two Nova Scotian watersheds: Tusket-Annis Rivers (hereafter "Tusket") and the Petite Rivière (Edge 1984). The Tusket population was reported to be anadromous, with adults migrating from salt water into fresh water during the autumn to spawn (Edge and Gilhen 2001). Some individuals were reported to overwinter within the lakes before returning to the sea in the spring. Reportedly once abundant, the Tusket population began declining in the 1940s and 1950s due to a combination of pressures including poaching as well as habitat degradation due to water acidification and the construction and operation of the Tusket hydro-electric facility (Edge and Gilhen 2001). Damming of this system also interfered with the annual migration of Atlantic Whitefish and evidence of a spawning run was not observed in the system following 1964 (Bradford et al. 2004). The last confirmed

Atlantic Whitefish capture within the Tusket occurred in 1982 and the species is now considered extirpated from the area (Edge 1984).

Within the Petite Rivière watershed, Atlantic Whitefish complete their lifecycle in three, small, connected lakes – Minamkeak, Milipsigate, and Hebb – which together span 16 km^2 in surface area (Bradford et al. 2004). A dam below Hebb Lake prevented return of anadromous fish from the ocean for nearly a century (Cook 2012); however, fish passage was added to the dam in 2012 (DFO 2018). Although the Petite Rivière population retains the ability to tolerate saline conditions, anadromy has not been observed since the ocean passage was restored. Atlantic Whitefish within this system face numerous threats including predation and competition by two aquatic invasive species (AIS) Smallmouth Bass (*Micropterus dolomieu*) and Chain Pickerel (*Esox niger*), which were first detected in the watershed in the mid-1990s (Bradford et al. 2004) and 2013 (Themelis et al. 2014), respectively. Declining habitat quality from surrounding land development, logging, and pollution resulting from urbanization pose threats as well (COSEWIC 2010). Though the population number has not been quantified, estimates of the effective population size are low at less than 100 individuals (Cook 2012).

Due to their evolutionary significance, 50% range decline from an already small range, and small population size, Atlantic Whitefish was the first fish species to be designated as endangered by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) in 1984 (Whitelaw et al. 2015). Their endangered status was formally recognized under the terms of the newly enacted federal Species at Risk Act in 2003 (DFO 2006). A breeding program was developed at the Mersey Biodiversity Facility by the Department of Fisheries and Oceans Canada (DFO) with operations running between 2000–2012 (Whitelaw et al. 2015). Between 2007–2009, approximately 12,000 juvenile Atlantic Whitefish were released into the Petite

Rivière watershed. Upwards of 4,000 Atlantic Whitefish juveniles and 7,000 larvae were also released into Anderson Lake, Dartmouth, NS, between 2005 and 2012 to assess the ability to establish new lake-resident populations from captively reared individuals (Whitelaw et al. 2015). The Anderson Lake trial was unsuccessful in establishing a new population (Bradford et al. 2015). Current conservation efforts for Atlantic Whitefish include annual removal of AIS from the lakes via electrofishing and destruction of Smallmouth Bass nests, annual juvenile counts beginning in 2015 (DFO 2018), and annual transport of early-stage juveniles to Dalhousie University for captive rearing beginning in 2018. At Dalhousie, a captive breeding program has begun, with successful reproduction first achieved in 2021. Release of some captively bred Atlantic Whitefish back into the Petite Rivière watershed occurred in summer 2022 with an additional set of releases occurring in 2023.

In 2018, a recovery strategy for the species was presented by the DFO which has the goals of (1) stabilizing the existing population, (2) expanding the range beyond the Petite Rivière watershed, and (3) restoring anadromy (DFO 2018). The numerous knowledge gaps surrounding Atlantic Whitefish biology and life history within the Petite Rivière watershed, including population size as well as spawning timing and habitat preference, present challenges for their conservation. The success of the DFO recovery goals will depend, in part, on our ability to detect Atlantic Whitefish in the wild for continued monitoring.

1.3 <u>Thesis outline</u>

In this thesis I describe development and use of a novel class of eDNA marker to increase the ability of eDNA analyses to detect low abundance species, using Atlantic Whitefish as a test case. I developed and validated a novel eDNA tool targeting *Sma*I-corII, a Coregoninae subfamily-specific short interspersed nuclear element (SINE), as well as a more conventional eDNA marker targeting the mitochondrial NADH dehydrogenase 4 subunit (ND4) using qPCR methods. The limit of detection and quantification of each marker type were assessed before the assays were tested on a transect of water samples ranging from 0 – 80 m away from a net pen housing juvenile Atlantic Whitefish in Milipsigate Lake. The assays were also run against water samples taken from tanks holding different numbers of Atlantic Whitefish to assess how eDNA yield scales with increasing fish densities (Chapter 2).

Next, I assessed the ability to detect genetic variation from eDNA samples amplified with *Sma*I-corII and DNA sequencing techniques. Consideration was taken to assess how widely available bioinformatic tools process data from a marker as variable as *Sma*I-corII before developing bioinformatic thresholds for *Sma*I-corII analysis. *Sma*I-corII sequence variants from Atlantic Whitefish eDNA samples were compared to Lake Whitefish eDNA samples to determine species-specific variants. Lastly, *Sma*I-corII variant dropout at decreasing eDNA concentrations was assessed through a dilution series of eDNA collected from captivity (Chapter 3).

In Chapter 4, I discuss the key findings of this thesis, summarize existing eDNA methods for detecting Atlantic Whitefish within their habitat, suggest future research avenues to expand upon the conclusions of this thesis, and highlight the significance of these findings as they related to other endangered freshwater fishes.

Chapter 2 – SINEs for the times: short interspersed nuclear elements increase ability to detect rare species via eDNA

2.1 Abstract

Accurate and sensitive methods are required for detection of species occurring at low abundances. While eDNA analysis has revolutionized aquatic monitoring, most studies to date have focused on targeting mitochondrial markers. Here, the ability to detect endangered Atlantic Whitefish via eDNA analysis using a novel nuclear DNA-based marker that targets a highly abundant nuclear transposable element, the *Sma*I-corII SINE, was compared to results obtained with a conventional mitochondrial marker that targets the ND4 subunit. In a field trial of water samples collected from a net pen housing juvenile Atlantic Whitefish and at varying distances from the net pen, the SINE marker provided greater detection sensitivity. Within the net pen, the abundance of SINE DNA copies was roughly 100 times greater than ND4. Outside of the net pen, the maximum distance of detection with the ND4 marker was 20 m whereas the SINE marker had detections up to 80 m, the maximum distance tested. An analysis of the relationship between eDNA yield and fish density using water from tanks holding whitefish at two densities revealed that differences in ND4-eDNA yield matched the 11-fold difference in fish densities, whereas the estimated difference in SINE-eDNA yield was only 6-fold. These results highlight the potential of eDNA markers that target high copy number nuclear DNA sequences for increased sensitivity of detection of aquatic species, but also point to the need for further work to better understand the relationship between eDNA yield and organismal abundance.

2.2 Introduction

The analysis of environmental DNA (eDNA) has revolutionized aquatic species monitoring by allowing the inference of species presence from genetic material shed into the environment (Barnes and Turner 2016). eDNA tools provide a non-invasive alternative to traditional methods, such as netting and electrofishing, which are typically more costly as well as time- and labour-intensive (Sakata et al. 2021). When employing eDNA tools, the DNA of a single species of interest can be targeted with species-specific primers, or "universal" primers can be used to detect the DNA of many species within the environmental sample to catalogue the community composition of an area ("metabarcoding"). Its low cost and multiple applications make eDNA a promising tool and it has been successfully used to detect endangered (Weltz et al. 2017) and invasive species (Rojahn et al. 2021), as well as to assess intraspecific variation of target species (Adams et al. 2019), and track species distributions pathways (Young et al. 2022). eDNA studies greatly benefit from *a priori* knowledge of target species habitat use and life history characteristics for accurate and efficient sampling design (Beng and Corlett 2020).

eDNA studies typically target mitochondrial DNA (mtDNA) due to its rapid evolution, allowing closely related species to be distinguished (Baillie et al. 2019), possibly slower degradation than nuclear DNA (nuDNA) (Foran 2006), and high copy number in the $10^1 – 10^3$ copies per cell range (Robin and Wong 1988). Despite its abundance, mtDNA copy number per cell is variable, influenced by cell type and body conditions (Minamoto et al. 2017). While mtDNA has an abundance advantage over single copy nuclear sequences, nuclear genomes also have many high-copy number sequences and some of these have potential utility as eDNA markers, as evidenced by the growing interest in incorporating multicopy ribosomal internal transcribed spacer (ITS) regions into eDNA analysis (Jo et al. 2019). Compared to mtDNA assays, targeting ITS markers led to greater sensitivity for detecting Bull Trout (*Salvelinus confluentus*) (Dysthe et al. 2018) and Common Carp (*Cyprinus carpio*) (Minamoto et al. 2017) eDNA owing to the higher copy number of the marker. During the spawning season of Macquarie Perch (*Macquaria australasica*), concentrations of the ITS marker were higher in eDNA samples than the mitochondrial 12S ribosomal RNA marker, though concentrations were found to be equal outside of the reproductive period (Bylemans et al. 2017). The quicker degradation of nuDNA may also lend itself to being more informative for determining the timing of species presence within the sampling area than mtDNA markers (Dysthe et al. 2018).

One class of highly abundant nuDNA repeat are transposable elements (TEs), dispersed repeats which move around the genome by creating new copies of themselves (Casacuberta and González 2013). TEs can be categorized into two classes: retrotransposons, which are derived from RNA and transposed to DNA via cDNA intermediates, and transposons, which are transposed from DNA to DNA (Hamada et al. 1997). Retrotransposons are the most abundant class of TEs, accumulating in the genome via a copy-and-paste amplification method (Elbarbary et al. 2016), and have been found across eukaryote species ranging from yeast to humans (Kramerov and Vassetzky 2011). One of three major subclasses of retrotransposons are short interspersed nuclear elements (SINEs). SINEs are non-coding and typically 100-500 base pairs long, often present in the range of 10^4 – 10^5 copies/cell (Hamada et al. 1997). A SINE-based assay developed for simultaneous detection of the dog (*Canis lupus familiaris*) SINEC_Cf element and human *Alu* Yb8 element from forensic samples was determined to be highly specific and sensitive, owing to the high copy number per cell of each SINE (Liang and Coyle 2020); however, SINE-based approaches have yet to be tested in eDNA applications.

Here I use Atlantic Whitefish (*Coregonus huntsmani*) as a test case for the development of a SINE-based eDNA marker. Atlantic Whitefish is an endangered species with a global distribution limited to three connected lakes comprising a total of 16km^2 in surface area (Bradford et al. 2004). Extreme low abundance and the legally protected status of the species effectively preclude detection of Atlantic Whitefish using traditional fish survey methods, and recovery efforts require alternative detection methods that are both maximally sensitive and non-invasive. The efficacy of the SINE eDNA marker was compared to results obtained with a 'conventional' eDNA marker targeting mtDNA. The SINE marker targeted *Sma*I-corII, a SINE found in the subfamily Coregoninae of the family Salmonidae (Hamada et al. 1997), and the mtDNA marker targeted a sequence within the mitochondrial NADH dehydrogenase 4 (ND4) subunit. The objectives of this research were to (i) determine whether the *Sma*I-corII eDNA marker enables more sensitive detection of Atlantic Whitefish eDNA compared to that obtained with the mtDNA marker, and (ii) test the species-specificity of the *Sma*I-corII eDNA marker. This work directly builds on previous studies comparing nuDNA and mtDNA environmental DNA (Bylemans et al. 2017; Minamoto et al. 2017; Dysthe et al. 2018) and represents the first exploration of a SINE-based eDNA marker.

2.3 <u>Methods</u>

2.3.1 Aquatron eDNA collection and processing

This study used adult Atlantic Whitefish raised and held in captivity at Dalhousie University's Aquatron Facility. Triplicate 1 L water samples were collected into sterile Nalgene bottles from a 2.3 m^3 tank housing 10 adult Atlantic Whitefish. A field blank was collected by exposing a sterile Nalgene bottle containing 1 L of distilled water to the air next to the tank for 10 s. Samples were immediately transported to the lab and filtered through two cellulose-nitrate filters with 5 μM (pre-filter) and 0.45 μM pore sizes using a peristaltic pump. Both filters were preserved in 100% ethanol and stored at -20°C until extraction. Prior to filtering samples, a 1 L filter blank was obtained by filtering distilled water through the peristaltic pump and preserving the filter in 100% ethanol.

Prior to extraction, filters were removed from the ethanol and placed into sterile Petri dishes and allowed to air dry for up to 48 hours. An extraction blank of 600 μL molecular grade water was included and treated the same as the filters. Once dry, filters were placed into 2mL tubes containing 900 μL of CTAB buffer (1.4 M NaCl, 2% w/v CTAB, 100 mM Tris pH 8.0, 10 mM EDTA, 1% w/v PVP) and 2 μL of proteinase K (20 mg/mL), then left to digest overnight at room temperature on a nutating mixer set to low. Following digestion, samples were centrifuged at 13,000 rpm for 1 min. Approximately 600 μL of supernatant was collected from each tube and placed into a new 1.5mL tube.

eDNA from filters was extracted using the phenol-chloroform-isoamyl alcohol (PCI) technique (Sambrook et al. 1989). Briefly, 600 μL of PCI (25:24:1 ratio) was added and samples were mixed on a vortex set to medium speed for 10 s before being centrifuged in a benchtop micro-centrifuge for 30 min at 13,000 rpm. Following this, 450 μL of the upper phase was collected and transferred to a new tube and 600 μL of chloroform was added. Samples were then mixed on a vortex for 10 s before being centrifuged for 5 min at 13,000 rpm. Afterwards, 400 μL of the upper phase was collected and transferred into a new tube. One-tenth volume of 3 M sodium acetate (40 μL) and two volumes of cold 100% ethanol (880 μL) were then added to the tube. Samples were manually mixed for 10 s and left to precipitate at -20°C overnight, then centrifuged at 13,000 rpm for 30 min. The 100% ethanol was decanted from each tube before 900 μL of 70% ethanol was added. The samples were centrifuged for a further 5 min at 13,000 rpm and the ethanol was then poured out of each tube. Tubes were left to air dry in a sterile bench drawer. The dry eDNA extracts were resuspended in 50 µL 1x TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) and stored at -20°C until analysis.

To limit contamination, all water sampling and filtering equipment were cleaned with 3% sodium hypochlorite for 10 min followed by a 10 min rinse with distilled water prior to use. All extractions took place in a dedicated eDNA lab which had been cleaned with 3% sodium hypochlorite and which had not previously housed PCR or genomic DNA (gDNA) products (Goldberg et al. 2016).

2.3.2 qPCR assay development

Hydrolysis probe quantitative PCR (qPCR) assays were designed for the mitochondrial ND4 subunit and the *Sma*I-corII SINE of Atlantic Whitefish. Primers and probes for each locus were designed using NCBI Primer-BLAST (see Appendix A, Table A1). Synthetic gBlock™ (IDT) fragments were designed for each marker by selecting a 500 bp segment that encompassed each of the forward and reverse priming regions as well as the probe. gBlocks arrived dried and were resuspended with 1x TE containing 10% tRNA. Following resuspension, the concentration of the *Sma*I-corII and ND4 gBlock stocks were 19.40 B and 9.75 B c/µL, respectively. To avoid possible contamination due to their high concentrations, the gBlock fragments were stored in the main Marine Gene Probe Lab (MGPL) rather than in the eDNA lab.

To ensure amplification success following assay development, the assays were run in a qPCR against both Atlantic Whitefish gDNA (1000 c/µL) and Aquatron-sourced eDNA. As eDNA samples came from a relatively small tank, they were assumed to be highly concentrated and were diluted 100x prior to the qPCR. Each well of the qPCR plate contained 5 µL 2X

PrimeTime™ Gene Expression Master Mix (IDT), 0.6 µL 10 µM primers, 0.15 µL 10 µM probe, 3.25 µL water, and 1 µL eDNA. qPCRs were run on a LightCycler® 480 System (Roche) and cycling parameters for each assay were 95°C 3 mins, 50x (94°C 15 s, 60°C 1 min, 72°C 15 s) 37°C hold.

Following amplification testing, the annealing temperature (T_A) of the assays was optimized by performing gradient PCRs against the Aquatron eDNA extracts (5 µM pre-filter) for each marker. Each gradient PCR contained 5 µL 2X Invitrogen™ Platinum™ SuperFi II Master Mix (Thermo Fisher Scientific), 0.5 µL 10 µM primers, 3.5 µL water, and 1 µL eDNA. PCR cycling parameters were: 98°C 2 min, 40x (98°C 15 s, T_A 30 s, 72°C 15 s), 72°C 5 min, 10°C hold. Tested annealing temperatures ranged from 60-75°C and 60-70°C for *Sma*I-corII and ND4, respectively. A no template control (NTC) was included on the plate for each gradient PCR. Amplification success was visualized on a 1% agarose gel. Following the gradient PCR, a 2-step PCR was performed on the same Aquatron eDNA extracts. The 2-step PCR used the same reaction mix but with the following cycling parameters: 2 min 98°C, 30x (98°C 15 s, 72°C 30 s), 72°C 5 min, 10°C hold. Following the 2-step PCR, the product was visualized on a 1% agarose gel to assess amplification success.

Assay specificity of *Sma*I-corII and ND4 were confirmed by performing qPCRs on gDNA (10 GE/µL) from co-occurring and related species: American Eel (*Anguilla rostrata*), Chain Pickerel (*Esox niger*), Smallmouth Bass (*Micropterus dolomieu*), Atlantic Salmon (*Salmo salar*), Brook Trout (*Salvelinus fontinalis*), Brown Trout (*Salmo trutta*), Lake Whitefish (*Coregonus clupeaformis*), Lake Trout (*Salvelinus namaycush*), Rainbow Trout (*Oncorhynchus mykiss*), and Atlantic Whitefish gDNA was included as a positive control. The qPCR was as described for the initial amplification test and the cycling parameters were: 95°C for 3 min, 50x (94°C 15 s, 70°C 1 min), 37°C hold.

2.3.3 LOD and LOQ assessment

The limit of detection (LOD) and limit of quantification (LOQ) for each marker were assessed following Klymus et al. (2020). The gBlock for each marker was diluted down to a 31,250 c/µL stock with 1x TE containing 10% tRNA (Klymus et al. 2020). Standards ranging from 128 to 2 c/µL were created from the 31,250 c/µL stock in a 2-fold dilution steps using the TE-tRNA solution. Standards were created one day prior to performing the qPCRs to adhere to contamination protocols by eliminating travel from the MGPL back into the eDNA lab. Standards were stored at -20°C overnight. All standards were created with low-bind tubes and tips.

The following day, for each assay, a qPCR was performed where each sample contained 5 µL 2X PrimeTime™ Gene Expression Master Mix (IDT), 0.6 µL 10 µM primers, 0.15 µL 10 µM probe, 3.25 µL water, and 1 µL eDNA. Each standard was run in 12 technical replicates and an NTC of molecular grade water was include on the plate. The NTC and reagents were laid into the plate in the eDNA lab before the plate was sealed and transferred to the MGPL where the standards were added. To limit the influence of evaporation and cross-contamination, the plate seal was cut and removed from the plate one row at a time with a clean single-edge razor blade when adding the standards. Standards were briefly vortexed and underwent a pulse spin in a microcentrifuge prior to plating 1 µL into each well. The seal was replaced before the next standard was plated.

The LOD of the assay was defined as the lowest concentration with amplification in >95% of technical replicates (12/12) while the LOQ was defined as the lowest concentration with less than 35% coefficient of variation (CV) between technical replicates (Klymus et al. 2020). CV was calculated with the following equation from (Forootan et al. 2017):

$$CV_{ln} = \sqrt{(1+E)^{(SD(Cq))^2 * \ln(1+E)} - 1}$$

where E represents the qPCR efficiency and SD(Cq) is the standard deviation of the replicate Cq values. qPCR cycling parameters for each assay were: 95°C for 3 min, 50x (94°C 15 s, 70°C 1 min), 37°C hold.

2.3.4 Field eDNA assessment: net pen-associated and density trials

In July 2022 triplicate 1 L water samples were collected within a 3 m^3 net pen with quarter-inch mesh size in Milipsigate Lake, NS in which 150 juvenile (~ 3 g) captive-bred Atlantic Whitefish were being held prior to release as part of a species recovery effort (J. Broome, DFO, pers. com. 2023). Samples were then collected from 5 – 80 m away from the pen, with the distance between samples doubling each time (Figure 1). At each distance, water samples were collected into two 500 mL polyethylene terephthalate (PET) bottles containing purchased drinking water. Prior to sampling, the water in the bottles was poured out ~1 m from the collection site before the bottle was submerged just beneath the water's surface to collect the sample. A 1 L field blank was collected at the net pen (0 m sample) by opening two 500 ml PET bottles and exposing them to the air for 10 s.

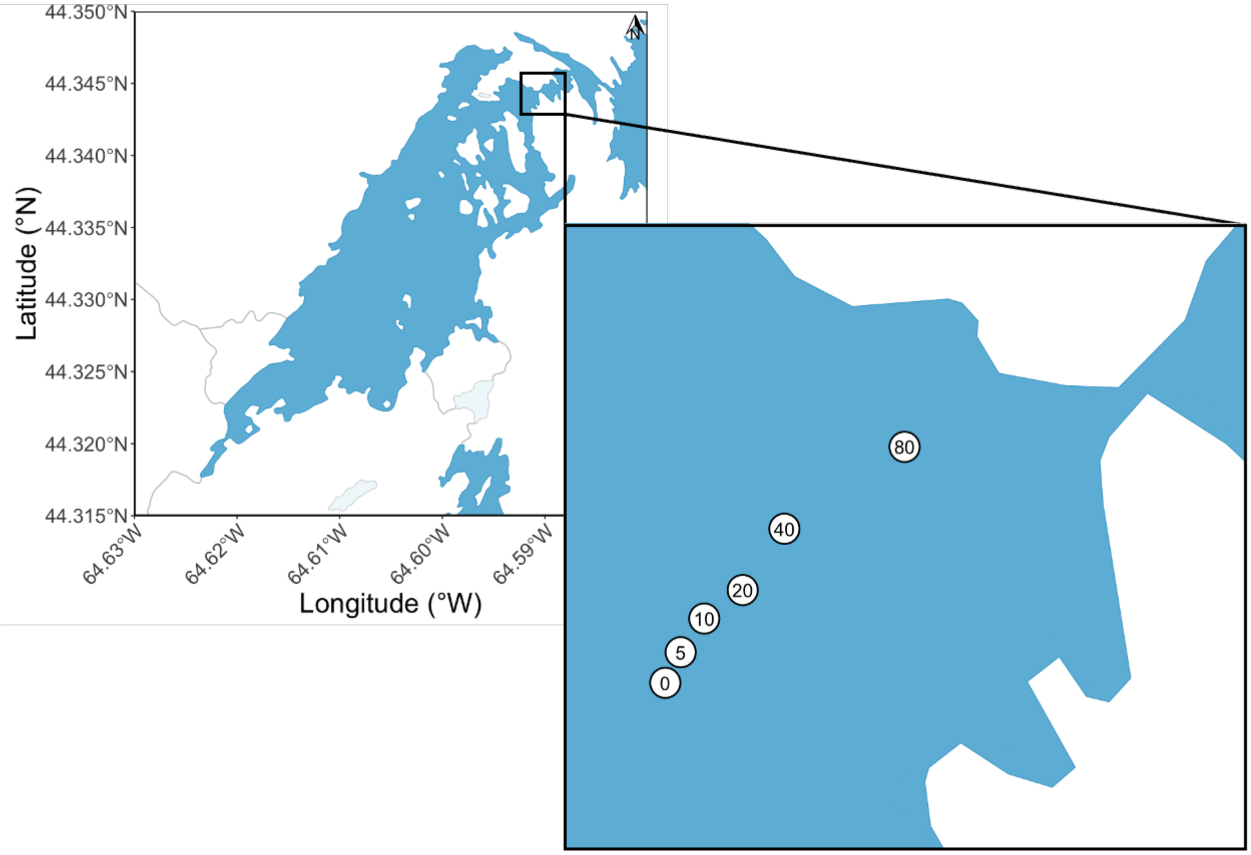

Figure 1 Net pen transect sampling locations in Milipsigate Lake, NS, Canada where the value inside the circle represents the distance (m) from the net pen.

Captive-bred juvenile Atlantic Whitefish held in ~500 L tanks with flow-through Petite Rivière water were used to assess how eDNA yield scales with increasing fish densities. A total of 50 juveniles were available for this trial and were split such that two tanks contained 2 fish each and two tanks contained 23 fish each (hereafter the "density trials'"). To account for any background Atlantic Whitefish eDNA in the Petite Rivière water the fish were being acclimated to, no fish were placed into a fifth (control) tank. Flow rate for the tanks averaged 14.45 ± 2.28 mL/s. Fish were placed into the tanks roughly 24 hours prior to eDNA sampling that followed same sampling procedure as described above. The 1 L field blank was collected next to the control tank. Welch's t-tests were performed in RStudio v.4.2.0 (R Core Team 2016) on the 2-

fish and 23-fish tanks for each assay to determine if the mean (μ) CP differed significantly (H_0: μ1 = μ2, H_A: μ2 ≠ μ1).

Water processing for each the net pen-associated and density trials followed the processes described in section 2.3.1. All extracts underwent a qPCR for each of the ND4 and *Sma*I-corII assays about two weeks after sample collection. Net pen-associated samples used the 5 μM filters whereas the density trials used the 0.45 μM filters. Each reaction well contained 5 μL 2X PrimeTime™ Gene Expression Master Mix (IDT), 0.6 μL 10 μM primers, 0.15 μL 10 μM probe, 3.25 μL water, and 1 μL eDNA. Cycling parameters for ND4 were as described in section 2.3.2 and for *Sma*I-corII were as described in section 2.3.3. As these qPCRs occurred prior to development and optimization of the gBlock standards, the concentrations of the samples were not determined in these assays.

Following development of the gBlocks, the net pen-associated and density trial samples were subjected to a second round of qPCRs for each assay that included gBlock standards. Standards ranging from 31,250 to 2 c/μL were used for the net pen-associated samples while the density trial samples used standards ranging from 6,250 to 2 c/μL. Standards were created using a 1x TE-tRNA (10%) solution as well as low-bind tubes and tips. Density trial samples underwent qPCRs for both the 0.45 μM and 5 μM filters. Each reaction well contained 5 μL 2X PrimeTime™ Gene Expression Master Mix (IDT), 0.6 μL 10 μM primers, 0.15 μL 10 μM probe, 3.25 μL water, and 1 μL eDNA. Cycling parameters were as described in section 2.3.3. Welch's t-tests were performed in RStudio v.4.2.0 on the 2022 and 2023 qPCRs to determine if the μ CP differed significantly between the years (H_0: μ1 = μ2, H_A: μ2 ≠ μ1).

2.4 Results

2.4.1 qPCR assay development, LOD and LOQ assessment

Following assay development, Atlantic Whitefish gDNA and Aquatron eDNA samples were amplified via qPCR with the *Sma*I-corII and ND4 markers to evaluate amplification success. Amplification was observed in each sample type for both markers (Table 1). For each sample type, the average CP number was lower for *Sma*I-corII assays than for ND4 with a difference of 7.57 and 6.36 cycles between the assays for gDNA and eDNA samples, respectively. These cycle differences correspond to 190x and 82x more *Sma*I-corII copies within the gDNA and eDNA samples than ND4, respectively.

Table 1 qPCR results of gDNA and eDNA samples amplified with SmaI-corII and ND4 assays following assay development.

	ND4		***Sma*I-corII**	
Sample Type	Replicates with amplification	CP (µ ± SD)	Replicates with amplification	CP (µ ± SD)
gDNA (1000 c/µL)	3/3	20.97 ± 0.06	3/3	13.40 ± 0.06
eDNA (100x dilution)	7/9	34.75 ± 0.96	9/9	28.39 ± 0.88

Visualization on an agarose gel revealed that amplification occurred at every annealing temperature tested for the ND4 and *Sma*I-corII assays. Likewise, agarose gel visualization confirmed amplification success for both markers in the 2-temperature PCR.

Both assays were tested against gDNA from related and co-occurring species to determine assay specificity. The qPCR was run for 50 cycles; however, to account for possible nonspecific amplification, a 40-cycle CP cut-off threshold was applied (Burns and Valdivia 2008). ND4 was found to be species-specific for Atlantic Whitefish, with no amplification observed in other species (Table 2). Following the qPCR with the *Sma*I-corII assay, amplification was observed in 3/3 technical replicates for Atlantic Whitefish and Lake Whitefish. A 1.86 cycle difference was observed between Atlantic Whitefish and Lake Whitefish extracts, corresponding to 3.6 times more *Sma*I-corII copies in Atlantic Whitefish than Lake Whitefish. The 7.63 cycle difference between Atlantic Whitefish gDNA extracts amplified with *Sma*I-corII and ND4 correspond to 198x more *Sma*I-corII target within the extract than ND4.

Table 2 qPCR results of specificity testing against related and co-occurring species for SmaI-corII and ND4 assays (CP =< 40 cycles).

		ND4		***Sma*I-corII**	
Common name	Scientific name	Replicates with amplification	CP (µ ± SD)	Replicates with amplification	CP (µ ± SD)
Atlantic Whitefish	*Coregonus huntsmani*	3/3	28.68 ± 0.19	3/3	21.05 ± 0.16
American Eel	*Anguilla rostrata*	0/3	-	0/3	-
Chain Pickerel	*Esox niger*	0/3	-	0/3	-
Smallmouth Bass	*Micropterus dolomieu*	0/3	-	0/3	-
Atlantic Salmon	*Salmo salar*	0/3	-	0/3	-
Brook Trout	*Salvelinus fontinalis*	0/3	-	0/3	-
Brown Trout	*Salmo trutta*	0/3	-	0/3	-
Lake Whitefish	*Coregonus clupeaformis*	0/3	-	3/3	22.91 ± 0.02
Lake Trout	*Salvelinus namaycush*	0/3	-	0/3	-
Rainbow Trout	*Oncorhynchus mykiss*	0/3	-	0/3	-

The LOD and LOQ of each assay were assessed via hydrolysis probe qPCR and a 40-cycle cut-off was applied to the obtained results (Table 3). Following the cut-off, no amplification was observed in the NTCs for either marker. For the *Sma*I-corII assay, the 8 c/µL standards had amplification in 11/12 replicates, corresponding to 92% replicates which is below the 95% recommendation of Klymus et al. (2020). Therefore, the LOD was determined to be 16 c/µL, with amplification in all technical replicates. When all samples which amplified were included in the standard curve calculation, the efficiency, error, and slope of the assay were of 2.21, 0.06, and -2.89, respectively and the LOQ was determined to be 64 c/µL (CV = 0.349). To improve the reliability of the standard curve (Zhang and Fang 2006), the 2 c/µL and 4 c/µL samples were removed from the calculation. Removing these samples decreased the efficiency, error, and slope of the assay to 2.05, 0.02, and -3.20, however the LOQ did not change from 64 c/µL (CV = 0.332). The LOD for the ND4 assay was four times lower, at 4 c/µL with amplification in all technical replicates. The LOQ for the ND4 assay was the same as for *Sma*I-corII at 64 c/µL (CV = 0.255). The efficiency, error, and slope of the assay were 1.98, 0.03, and -3.36, respectively.

Table 3 Results of SmaI-corII and ND4 qPCR assays used to calculate LOD and LOQ. Bolded-underlined values indicate the LOD (defined as >95% of technical replicates with amplification) and bolded values indicate the LOQ (defined as <35% variance among technical replicates). LOD and LOQ assessment followed guidelines from Klymus et al. (2020). Samples which amplified but were excluded from the standard curve calculation are indicated by an asterisk (*).

	ND4	***Sma*I-corII**
gBlock concentration (c/µL)	Replicates with amplification	Replicates with amplification
2	6/12	4/12*
4	**<u>12/12</u>**	7/12*
8	12/12	11/12
16	12/12	**<u>12/12</u>**
32	12/12	12/12
64	**12/12**	**12/12**
128	12/12	12/12
NTC	0/12	0/12

2.4.2 Field eDNA assessment: net pen-associated samples

To assess the distance of detection (in metres) of each marker, samples were collected at distances up to 80 m from a net pen in Milipsigate Lake that housed 150 ~3g juvenile Atlantic Whitefish. One PCR replicate of the field blank had amplification with a CP of 25.77 in the *Sma*I-corII assay while no blanks amplified with ND4. Each marker amplified in all nine replicates within the net pen (0 m) with average CPs differing by 6.77 cycles between *Sma*I-corII and ND4 assays with CPs of 24.80 and 31.57, respectively, indicating a greater abundance of *Sma*I-corII than ND4 sequences within the pen by over 100-fold (Table 4). Following a 40-cycle cut-off, ND4 had a detection at the 5, 10, 20 m distances in one replicate at each distance and no detection at 40 or 80 m (Figure 2). In contrast, *Sma*I-corII had detection at all distances out to 80 m, and at each distance detections occurred in at least one-third of replicates. An outlier with a CP 1.5 standard deviations less than the mean (CP = 16.24), was observed in a single technical replicate at 10 m for *Sma*I-corII. The outlier was removed from the dataset prior to any statistical testing. As the *Sma*I-corII marker had detection in the furthest distance sampled, the distance of detection was not determined. The distance of detection for ND4 was determined to be 20 m.

Table 4 Mean crossing point (CP =< 40 cycles) and number of technical replicates with detection of the net pen-associated samples for SmaI-corII and ND4 assays. The number of technical replicates reported refers to the number of replicates which had amplification minus any which were inferred to be outliers. Samples where outliers were removed are identified with an asterisk (*).

	ND4		***Sma*I-corII**	
Distance	Replicates with amplification	CP (µ ± SD)	Replicates with amplification	CP (µ ± SD)
Blank	0/3	-	1/3	25.77
0	9/9	31.57 (±0.83)	9/9	24.80 (±0.77)
5	1/9	36.6	5/9	35.87 (±1.94)
10	1/9	36.2	3/9*	37.97 (±1.17)
20	1/9	37.53	7/9	35.44 (±0.83)
40	-	-	3/9	38.47 (±1.44)
80	-	-	5/9	38.26 (±1.21)

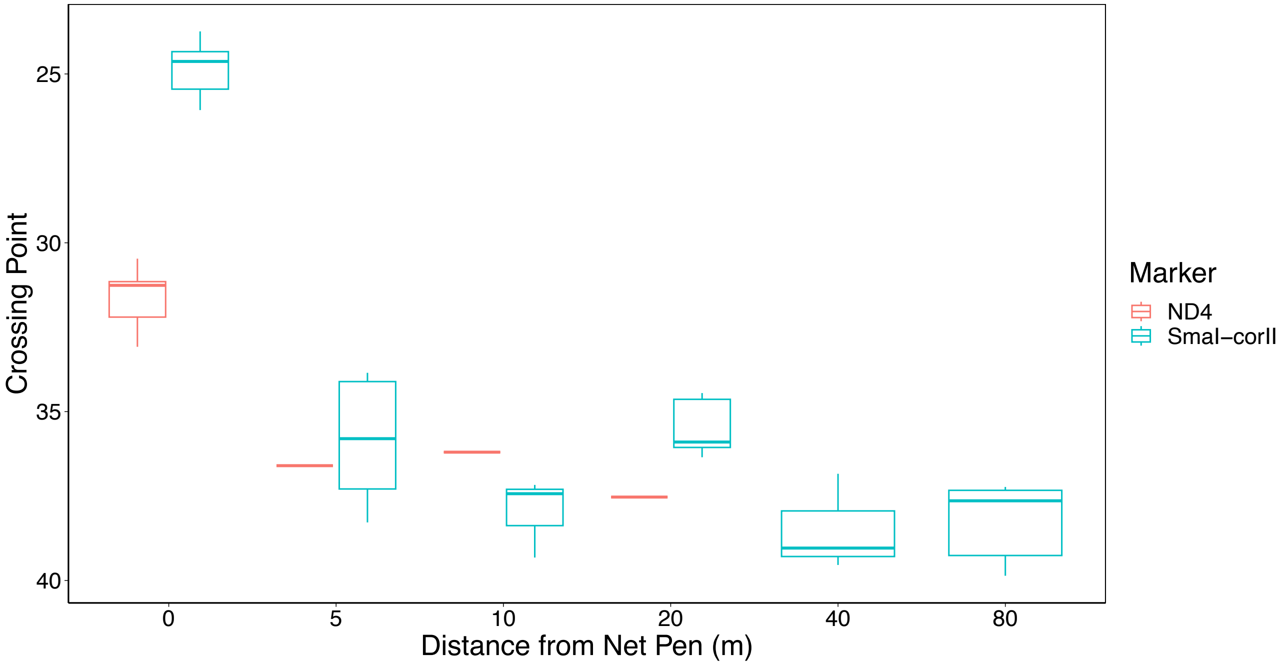

Figure 2 Crossing points (CP =< 40) cycles of eDNA samples amplified with SmaI-corII and ND4 assays taken up to 80 m away from a net pen housing 150 ~3g juvenile Atlantic Whitefish in Milipsigate Lake, NS.

2.4.3 Field eDNA assessment: density trial samples

To assess how eDNA detection scales with increasing fish densities, samples taken from tanks containing 2 and 23 ~3g juvenile Atlantic Whitefish were amplified with *Sma*I-corII and ND4. Amplification was observed with *Sma*I-corII in one PCR replicate of the field blank collected beside the control tank (CP = 38.31) however no amplification was observed from samples collected from within the control tank with either marker. An outlier with a CP 2.5 standard deviations less than the mean (CP = 16.29) was observed in a single technical replicate in tank 3 for the *Sma*I-corII assay. The outlier was removed from the dataset prior to any statistical testing. *Sma*I-corII assay had detections in more technical replicates and lower CPs than when samples were analyzed with ND4 (Table 5). For tanks housing 2 fish, *Sma*I-corII detected between 51 – 95x more target molecules than ND4 while in the 23-fish tanks the mean cycle difference between the assays corresponded to 12 – 108x more *Sma*I-corII target than ND4.

Table 5 Average crossing point (CP) and number of technical replicates with detection for samples taken from tanks holding 2 or 23 ~3g juvenile Atlantic Whitefish and amplified with SmaI-corII and ND4 assays. The number of technical replicates reported refers to the number of replicates which had amplification minus any which were inferred to be outliers. Samples where outliers were removed are identified with an asterisk (*).

		ND4		***Sma*I-corII**	
Tank	Number of fish	Replicates with amplification	CP (μ ± SD)	Replicates with amplification	CP (μ ± SD)
Blank	-	0/3	-	1/3	38.31
1	0	0/9	-	0/9	-
2	2	8/9	34.52 (±0.98)	9/9	27.95 (±0.57)
3	2	7/9	34.48 (±1.11)	7/9*	28.81 (±0.17)
4	23	9/9	32.90 (±0.87)	9/9	26.15 (±1.35)
5	23	8/9	28.96 (±0.72)	9/9	25.43 (±1.15)

For the *Sma*I-corII assay, there was a significant difference in the mean CP of tank 2 and tank 3 (t(9.83) = -4.27, p = 0.002, Welch's t-test). No significant difference was observed between the mean CPs of tanks 4 and 5 (t(15.58) = 1.21, p = 0.244, Welch's t-test). The opposite pattern was observed in the ND4 assay, where the difference in mean CP of tanks 2 and 3 was not significant (t(12.15) = 0.08, p = 0.941, Welch's t-test) while the difference between tanks 4 and 5 was significant (t(14.93) = 10.25, p = 3.786e-08, Welch's t-test).

Consolidating the tanks by number of fish, for each assay there was greater variance across the CPs of the 23-fish tanks than the 2-fish tanks (Figure 3). When amplified with *Sma*I-corII, the average CP for the 2-fish tanks was 2.53 cycles greater than the 23-fish tanks (2-fish = 28.32 ±0.62, 23-fish = 25.79 ±1.27). As each cycle of a qPCR represents a doubling of DNA, this cycle difference corresponds to approximately 6 times more target DNA in the tanks with 23 fish compared to 2. For the ND4 assay, the difference between the average CP for the 2-fish and 23-fish tanks was larger than for *Sma*I-corII, 3.46 cycles (2-fish = 34.50 ±1.00, 23-fish = 31.05 ±2.17). This difference represents approximately 11 times more target DNA in the 23-fish tanks than the 2-fish tanks.

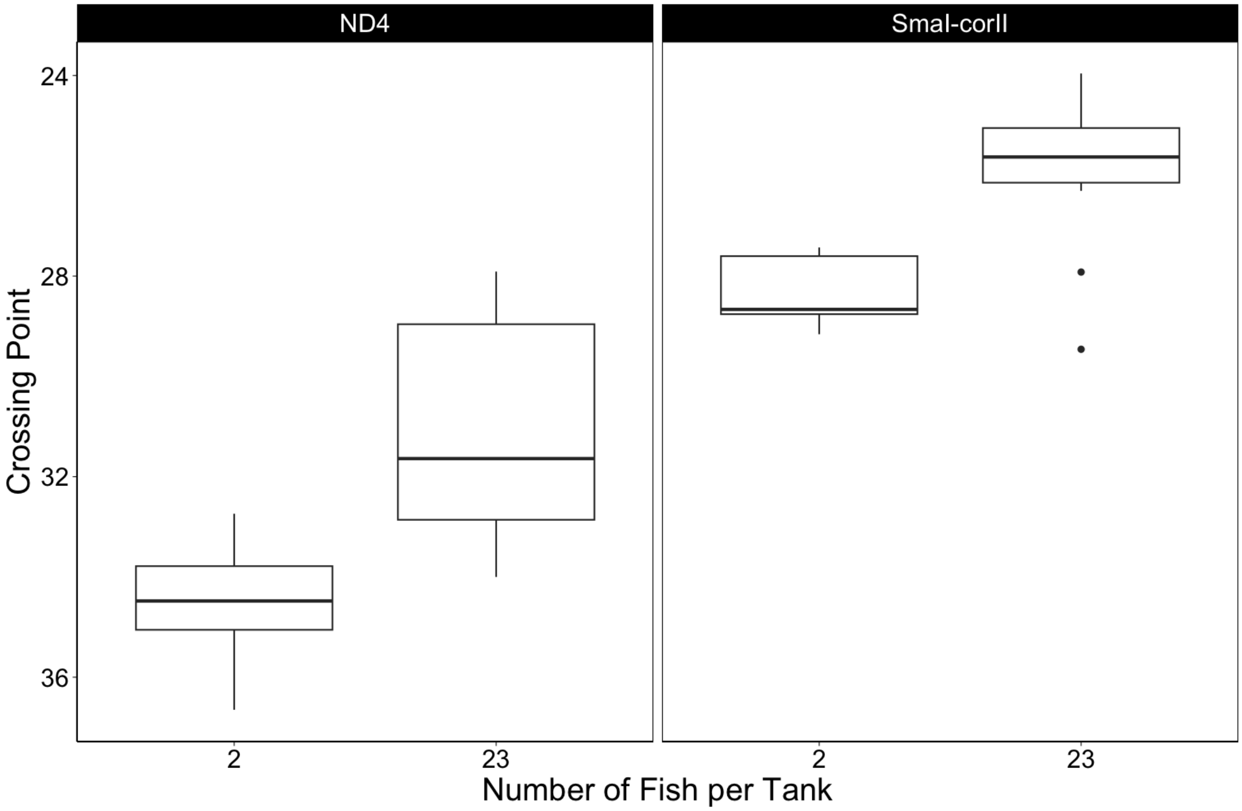

Figure 3 Crossing points of eDNA samples amplified with SmaI-corII and ND4 assays taken from tanks holding 2 or 23 ~3g juvenile Atlantic Whitefish.

2.5 Discussion

The continued decline of aquatic populations (Dudgeon 2010) warrants sensitive, non-invasive detection methods to inform conservation and management measures, such as eDNA tools (Abbott et al. 2021). The presence and subsequent detection of environmental DNA has been shown to serve as a valuable proxy for species presence and particularly advantageous for detecting rare or cryptic species compared to traditional methods (Sigsgaard et al. 2015). However, challenges remain regarding the sensitivity of eDNA to assess species occurring at extremely low abundances where the approach is likely to most valuable. Here, I demonstrated a SINE-based eDNA marker targeting *Sma*I-corII lead to an increased ability for detecting Atlantic Whitefish eDNA in the field compared to a conventional eDNA marker targeting the mitochondrial ND4 subunit assessed via qPCR. Additionally, I assessed how the eDNA yield of the two markers scaled with increasing fish densities to determine if yield was proportional to density and could be related to abundance estimates.

2.5.1 Assay development

*Sma*I-corII and ND4 assays were each able to amplify Atlantic Whitefish DNA from gDNA and eDNA samples. Across the assays for each sample type, the difference in mean CPs corresponded to 190x and 82x more *Sma*I-corII target within the gDNA and eDNA samples than ND4, respectively. As this initial qPCR was performed to confirm the ability of the assay to detect its target marker within different samples types, starting concentration of the eDNA samples was not quantified prior to testing nor were concentrations determined during the qPCR. The standard deviation of each assay was greater in the eDNA samples than the gDNA, possibly owing to the heterogenous nature of eDNA (Beng and Corlett 2020).

The specificity of each assay was tested against gDNA from co-occurring and related species and the ND4 assay was determined to be species-specific, with no amplification observed in the other species assessed. The ND4 assay produced a mean CP value for Atlantic Whitefish which was 7.63 cycles greater than the CP of the *Sma*I-corII marker. This difference corresponds to nearly 200 times more copies of *Sma*I-corII in the Atlantic Whitefish extract compared to ND4 which underscores the marker's advantage, as it occurs in greater quantities within a given sample. Using the *Sma*I-corII assay, amplification was observed in all PCR replicates for both Atlantic Whitefish and Lake Whitefish but not in any other species, confirming the specificity of this marker as Coregoninae subfamily-specific (Hamada et al. 1997). As Atlantic Whitefish extracts had an earlier crossing point in the qPCR than the Lake Whitefish samples at equal starting concentrations (10 GE/µL), it can be inferred that Atlantic Whitefish have ~3 times as many *Sma*I-corII copies across their genome than Lake Whitefish. As the oldest member of the Coregonus genus, Atlantic Whitefish represent a distinct lineage with estimates of their divergence from their common ancestor with other extant *Coregonus* species put around 14 million years ago (MYA) compared to the more recent divergence of Lake Whitefish, estimated at less than 5 MYA (Crête-Lafrenière et al. 2012). Though TE copy number is not directly correlated to the evolutionary age of a species (Zhang et al. 2023) and TE copy number has been observed varying between populations of the same species (Biémont and Vieira 2006), the large evolutionary distance between Atlantic Whitefish and Lake Whitefish may have contributed the observed difference in SINE copy number between the species. For most organisms, TE content is correlated to genome size (Wells and Feschotte 2020), and analysis of the complete Atlantic Whitefish genome would allow for direct comparisons to the Lake Whitefish genome (2.68 GB – 2.76 GB; Mérot et al. 2023) to provide further insights into how *Sma*I-corII copy number relates to genome size for these two species.

When assessing the LOD and LOQ of the assays on synthetic DNA, the two assays had equal LOQs of 64 c/µL; however, the LOD of the ND4 assay (4 c/µL) was four times lower than that of *Sma*I-corII (16 c/µL), indicating the ND4 assay is more efficient at detecting target molecules within a sample. Redesigning the *Sma*I-corII primers may increase the sensitivity of this assay, however further work would be required to confirm if redesigned primers yield a lower LOD. To obtain efficiency and slope values within an acceptable reporting range for the *Sma*I-corII assay during LOD and LOQ determination, the 2 and 4 c/µL were removed from the standard curve calculation due to an initial efficiency value of 2.21. This value corresponded to the PCR product increasing by 2.2 times each cycle, 20% more than would be expected if the assay had 100% efficiency. Inflated efficiencies can be due to inhibitors present within the sample; however, these are typically more prevalent in highly concentrated samples (Svec et al. 2015) and were not expected in the gBlock fragments used during testing. To ensure the reliability of standard curves, efficiencies should fall between 80-115% with a slope between -3.0 and -3.9 and have an R^2 value of 0.95 (Zhang and Fang 2006) which corresponds to an error of less than 0.05 for this study. Removing the 2 and 4 c/µL samples, all of which were below the determined LOD, resulted in the standard curve falling within this range. Notably, the LOD determined in this study used a more conservative threshold of 100% detection than the 95% recommended by Klymus et al. (2020) due to plate layout constraints.

Assay development could benefit from rerunning the LOD/LOQ assessment using a higher number of technical replicates for better resolution. Amplification success of *Sma*I-corII and ND4 following a 2-step PCR indicates that in future work, the time required to run the assays can be reduced by combining the annealing and extension steps. Cycle cut-offs are arbitrary (Burns and Valdivia 2008) and a 40-cycle one was applied to the qPCRs run during assay development; however, when these assays are applied to field samples, a lower cycle number should be used during the qPCR to avoid late, nonspecific amplification which increases with increasing cycle numbers (Cha and Thilly 1993). Determination of an optimal cycle number will require further analysis.

For the purposes of Atlantic Whitefish identification in the wild using the SINE maker, the lower specificity of *Sma*I-corII should not pose a problem for interpretation of eDNA results as Lake Whitefish are not known to occur within the Petite Rivière. If Atlantic Whitefish are to be introduced beyond their limited native range as dictated in their federal recovery plan (DFO 2018), care will need to be taken with interpreting eDNA results when using the *Sma*I-corII assay as this marker provides subfamily-level resolution, however potential translocation sites are pre-screened for the presence of Lake Whitefish and the species is only known to occur in a dozen or so lakes around the province (Bradford and Mahaney 2004).

2.5.2 Field eDNA assessment: net pen-associated samples

Each assay was found capable of detecting Atlantic Whitefish eDNA from captivity, however these tools were developed for conservation applications and as eDNA is expected to be influenced by the environment it occurs in (Barnes and Turner 2016), the assays needed to be further validated against water characteristic of Atlantic Whitefish's natural environment. When the assays were tested on water samples collected within a net pen housing 150 juvenile Atlantic Whitefish and up to 80 m away, the *Sma*I-corII assay had increased detection abilities compared to ND4 with detection at every sampled distance. Amplification was observed within the field blank of the *Sma*I-corII assay; however, as it occurred in only one PCR replicate and returned a CP similar to the 0 m samples, the contamination likely occurred during plate layout as the samples were plated concurrently and adjacent to one another. The observed outlier in the 10 m samples for the *Sma*I-corII assay was likely due to a pipetting error.

Assessing the net pen-associated samples with ND4 resulted in detection up to 20 m away from the net pen. Though 20 m is four times less than the maximum distance with detection observed using the *Sma*I-corII marker, this distance of detection is greater than what was observed by (Brys et al. 2021) in an assessment of eDNA transport of seven fish species in a lentic system via metabarcoding, in which the maximum distance of detection ranged from 5 – 10 m depending on the species. A greater distance of detection was observed by Dunker et al. (2016) where eDNA from cages housing one or two Northern Pike (*Esox lucius*) in four lentic systems was detected up to 40 m away. Across the 1, 10, and 40 m samples, the probability of eDNA detection decreased with distance as represented by a decrease in DNA copy number (Dunker et al. 2016), a trend not observed in this study as the ND4 CP was similar at the 5, 10, and 20 m distances and each only had detection in 1/9 technical replicates. Notably, the concentration of eDNA within the 0 m distance varied between studies which may partly explain the observed differences in distance of detection.

Across the distances sampled, no discernible pattern was observed regarding the mean CP or number of replicates for detection with the *Sma*I-corII marker. This homogenous result is similar to findings by Wood et al. (2020) who demonstrated that in lotic systems, Atlantic Salmon eDNA originated from the source sentinel cage housing 3 – 63 juveniles as a plume which was followed by even dispersal through the midstream before accumulating in stream margins further downstream (>1000 m). Although flow in a lentic system is expected to be lower than in lotic systems and most eDNA transport studies to date have concentrated on lotic systems (Brys et al. 2021), a plume-like eDNA cloud originating from the net pen may explain why CP varied by such a small amount moving away from the pen as samples were collected toward the lake outflow. Additionally, the observed eDNA dispersal could have been partly attributed to wind-facilitated water flow (Zhang et al. 2020) a mechanism reported to affect surface water mixing of small lakes (George and Edwards 1976).

Factors influencing the fate of eDNA such as surface water flow patterns, circulation dynamics, habitat heterogeneity, microbial community, and pH are expected to vary among aquatic systems (Zhang et al. 2020). Further analysis should be concentrated on understanding the specifics of Atlantic Whitefish eDNA dynamics within the Petite Rivière system to better relate eDNA detections to the species physical location within the waterbody for robust and accurate interpretation of eDNA results.

2.5.3 Field eDNA assessment: density trials

Samples taken from tanks holding 2 or 23 juvenile Atlantic Whitefish were assessed to determine how eDNA yield scales with increasing fish densities. Previous studies have reported eDNA concentration is positively correlated to species abundance (Klobucar et al. 2017) and biomass (Takahara et al. 2012), however the numerous environmental factors affecting eDNA degradation reduce this correlation in field settings compared to controlled settings (Rojahn et al. 2021). Here, I demonstrated that the relationship between density and eDNA yield also differs between marker types. With a CP difference between the 2-fish and 23-fish tanks corresponding to an expectation of 11 times more eDNA in the 23-fish tank, ND4 scaled closely with this predicted ration, while *Sma*I-corII did not, with an inferred difference of only 6 times the eDNA in the 23-fish tanks compared to the 2-fish tanks.

Many studies have assessed influences on eDNA shedding rate (Maruyama et al. 2014; Klymus et al. 2015; Sassoubre et al. 2016; Sansom and Sassoubre 2017; Nevers et al. 2018) but fewer have examined the differences between nuclear-eDNA (nu-eDNA) and mitochondrial-eDNA (mt-eDNA). For Japanese Jack Mackerel (*Trachurus japonicus*), nu-eDNA (ITS1) shedding rates were observed to increase with increased biomass and temperatures, however this trend was generally echoed in the mt-eDNA analysis (cytochrome b) as well (Jo et al. 2019). The ratio between mt-eDNA and nu-eDNA concentration was shown to decrease with larger fish biomass (Jo et al. 2019), a trend not observed within this study.

Differences in the ratio between the nu-eDNA (ITS1) and mt-eDNA (12S) were also observed to differ in response to reproductive activity of Macquarie Perch (*Macquaria australasica*), owing to their broadcast spawning strategy where spermatozoa, which contain relatively low amounts of mtDNA and well protected nuDNA, is released into the water column (Bylemans et al. 2017). Interestingly and unlike the results in this study, Bylemans et al. (2017) did not observe a difference of the two eDNA markers outside of the spawning period. Notably, this study used a small sample size of juvenile Atlantic Whitefish and eDNA shedding has previously been observed differing between life stages; Maruyama et al. (2014) demonstrated that when normalized by biomass, Bluegill Sunfish (*Lepomis macrochirus*) juveniles shed 4x more mt-eDNA than adults; however, when normalized by number of individuals, adults shed 12x more than juveniles. These results highlight the challenges of relating eDNA to abundance and further work should be done to examine the different factors affecting Atlantic Whitefish *Sma*I-corII eDNA shedding and decay to better contextualize eDNA detection results.

2.5.4 Limitations and challenges

The net pen-associated and tank-density Atlantic Whitefish were part of a concurrent but unassociated project and the samples used within this study were collected opportunistically. As such, the sample size for the density trial was small at only 50 fish across two treatments.

Additionally, as a result of the opportunistic nature of this sampling, samples were collected before the assays were fully developed. Though qPCRs were run within two weeks of collection, they could not be used to quantify the concentration as issues with the standards were not yet resolved. Extracts were stored at -20°C following extraction while the issues with the qPCR assays were addressed. The subsequent qPCR occurred a year following the first one and although concentrations were determined for the net pen-associated and density trial samples using both *Sma*I-corII and ND4 assays (see Appendices B and C for comparisons of 2022 and 2023 qPCRs), significant differences in mean CP between the years were observed for both sample sets. Therefore, the determined concentrations are not believed to be reflective of the concentration of the samples at the time of collection due to the discrepancies between the 2022 and 2023 results. These discrepancies indicate degradation of the eDNA extracts occurred. Possible explanations for the observed degradation include the multiple freeze-rethaw cycles the samples were subjected to during assay validation, inadequacy of the -20°C storage temperature to preserve eDNA extracts, and the loss of power for three days following Post-Tropical Storm Fiona in 2022, which caused the samples to reach room temperature for more than a day. The opportunistic nature of the field samples meant resampling was not possible. The observed degradation supports timely analysis of samples to ensure results are reflective of the environment at the time of sampling. Future work should examine the effect of storage method and temperature as well as freeze-rethaw cycles on eDNA extract quality to assess the feasibility of reanalysis of extracts following sampling.

Development and validation of eDNA tools greatly benefit from *in situ* testing and many of the proposed next research steps regarding Atlantic Whitefish eDNA tools require such analysis. The non-invasive nature of eDNA tools make them particularly advantageous for endangered species where traditional sampling methods can result in undue harm, however gaining access to endangered species during assay development and validation can be challenging and timely. Due to its endangered status, gaining direct access to Atlantic Whitefish for experimental purposes, such as for the proposed eDNA shedding and decay analyses, required timely approval processes and the timeline of this study did not support such an approach. Future work should budget to undertake the bureaucratic approval processes as determining Atlantic Whitefish eDNA shedding and decay rates will strengthen our understanding of any eDNA results obtained from the markers developed within in study.

Lastly, the net pen-associated and density trial samples used difference filter pore sizes (see Appendix D for a pore size comparison of the density trial samples in 2023 using *Sma*I-corII). Analysis of these samples occurred during concurrent analysis of lab protocols, including analysis of filter pore size choice on eDNA yield. Due to the turbid and tannic nature of the Petite Rivière system, a filter/pre-filter system had been chosen to limit filter clogging with an expectation that the 5 µM pre-filter would capture larger particles and the 0.45 µM filter would collect the eDNA (Li et al. 2018). The observed degradation resulted in the samples being unable to be reanalyzed. A significant difference was observed between the pore sizes, with the 5 µM filter retaining higher eDNA yield than the 0.45 µM (Table D1). As such, the CPs of the density trial samples are likely over-estimated.

2.5.5 Conclusions

Comparisons of results obtained with a novel SINE marker with those obtained with a conventional mtDNA marker have revealed the SINE-based approach enables more sensitive detection ability of Atlantic Whitefish eDNA. Though the LOD was lower for ND4, indicating this assay is more efficient at detecting target molecules within a sample, the net-pen associated samples highlight the increased ability to detect *Sma*I-corII copies in a field setting owing to its increased abundance within the genome. Within the net pen, there were roughly 100x more *Sma*I-corII copies present than copies of ND4, as demonstrated by the large cycle difference between the assays. Furthermore, *Sma*I-corII enabled detection of Atlantic Whitefish DNA up to 80 m away from the net pen, compared to the 20 m distance of detection with the mtDNA marker. Evaluating the relationship between eDNA yield and increasing fish densities revealed a noteworthy aspect in detecting a SINE repeat through eDNA analyses that requires further investigation. Specifically, the *Sma*I-corII yield exhibited a 6-fold increase when fish densities rose by 11-fold. In combination, these results suggest SINEs offer increased sensitivity of detection of aquatic species, but more work is needed to determine the potential to move beyond detection/non-detection inferences. Though the field samples were unable to be quantified due to degradation, the observed sample degradation highlights the importance of ground truthing eDNA lab practices and protocols prior to use.

Targeting a highly abundant nuclear repeat such as *Sma*I-corII in eDNA analysis will aid ongoing Atlantic Whitefish conservation plans by increasing sensitivity of detection in monitoring within their last remaining natural habitat. Moreover, SINEs are present in most eukaryotes (Kramerov and Vassetzky 2005) and are particularly well studied within salmonids, a family containing many commercially and culturally important species (Matveev and Okada 2009). As biodiversity continues to decline and species of interest become rarer, SINEs and other repetitive nuclear elements can provide a more sensitive alterative to mitochondrial markers.

Chapter 3 – Evaluating the potential of SINEs to assess genetic diversity from eDNA samples

3.1 Abstract

Environmental (e)DNA has become well established as an aquatic biodiversity monitoring tool. In the face of biodiversity decline, there is growing interest in expanding the scope of eDNA analysis to include information about intraspecific diversity. Previous studies have demonstrated the utility of mitochondrial haplotypes and nuclear microsatellite alleles detected from eDNA samples for population genetics. Here, Atlantic Whitefish were used as a test case to evaluate genetic diversity of a high copy number nuclear marker (SINE, *Sma*I-corII) from eDNA samples. Genomic DNA (gDNA) from 16 Atlantic Whitefish was used to compare two bioinformatic pipelines. The *Sma*I-corII consensus sequence and four additional ASVs were detected within the gDNA samples. Environmental water samples containing Atlantic Whitefish DNA were then analyzed and all five ASVs were detected. A total of six ASVs were detected in eDNA samples from a lake known to contain Lake Whitefish. Four ASVs were detected in both species, one ASVs was unique to Atlantic Whitefish, and two ASVs were unique to Lake Whitefish. These results indicate SINEs can be used to assess genetic diversity via eDNA analysis, although commonly available bioinformatic tools must be optimized for SINE ASV detection. Overall, transposable elements have the potential to provide population-level insights. Future work should examine the utility of this approach in more abundant species that are likely to harbour more genetic variation.

3.2 Introduction

Detection of species presence and estimation of abundance are common applications of environmental (e)DNA tools, but there is growing interesting in determining what other information can be reliably gathered from eDNA data, such as analysis of genetic diversity which is a critical component of a species' ability to adapt, survive, and contribute to the overall health and stability of ecosystems (Hoban et al. 2021). As with detection-based applications of eDNA, most eDNA genetic diversity studies to date have targeted mitochondrial DNA (mtDNA) (Sigsgaard et al. 2020). Mitochondrial haplotypes have been detected from eDNA samples using quantitative (q)PCR (Uchii et al. 2016; Goricki et al. 2017), droplet digital PCR (Baker et al. 2018), and DNA sequencing techniques (Sigsgaard et al. 2016; Stat et al. 2017; Parsons et al. 2018; Marshall and Stepien 2019; Stepien et al. 2019; Turon et al. 2020). Advantages of mtDNA markers over nuclear DNA (nuDNA) include its high copy number per cell (Robin and Wong 1988) and possibly slower degradation than nuDNA (Foran 2006). Furthermore, due to the long history of using mtDNA in population genetics and DNA barcoding, many reference databases of mtDNA sequences exist (Sigsgaard et al. 2020).

The advantages of mtDNA over nuDNA are not without trade-offs; as mtDNA is usually maternally inherited, it will only provide partial insights into population differentiation (Sigsgaard et al. 2020). To counter this disadvantage, nuDNA markers are increasingly being used in eDNA analyses, echoing an earlier shift to nuDNA markers in population genetic studies (Andres et al. 2021). Population-level analyses of Round Goby (*Neogobius melanostomus*) using a panel of 28 microsatellite loci found allele frequencies estimated from eDNA were close to those obtained from genotyped tissue samples; however, several low-frequency alleles were not recovered in the eDNA samples (Andres et al. 2021). Moreover, Jensen et al. (2020) found that nuclear introns, non-coding sections of DNA anticipated to exhibit high variability due to limited functional constraints, of Whale Sharks (*Rhincodon typus*) retrieved through eDNA analysis were insufficient for making reliable population-level inferences using DNA sequencing methods; of 12,411 variants initially detected, half were estimated to have resulted from PCR/sequencing errors. Increasing the filtering thresholds from 1–2x coverage to 10x decreased the number of variants to 22. These studies highlight that while nuclear variation can be detected from eDNA samples, challenges remain in discerning between rare variants and erroneous sequences. To offset the abundance disadvantage of nuclear eDNA (nu-eDNA) markers, PCR assays may be multiplexed to target multiple microsatellites or single nucleotide polymorphisms (SNPs) in a single reaction. Amplifying multiple nuclear loci may stochastically result in an adequate subset of markers being detected, thereby increasing the likelihood of detection by targeting multiple markers rather than just one. This approach requires careful optimization as multiplexed primers may interact with one another and cause inhibition, a challenge which may be heightened by the variable and uncontrollable starting concentration of eDNA samples (Andres et al. 2023a).

One type of nuclear marker which has to-date been overlooked in eDNA analyses are transposable elements (TEs). TEs are highly repetitive mobile elements located throughout the genome, accounting for nearly 50% of the genome in mammals (Lander et al. 2001), 7 – 56% in some teleost genomes (Gao et al. 2016), and up to 80% in some plant species (SanMiguel et al. 1996). TEs comprise two classes: those which require an RNA-intermediate and reverse transcriptase for replication (retrotransposons) and those which do not (transposons). Additionally, TEs are categorized by their ability to transpose. Autonomous TEs can transpose on their own while those which require another TE to transpose are nonautonomous (Stapley et al. 2015). The diversity of TEs differs among major taxa; for example, ray-finned fishes and amphibians have more diverse TEs within their genomes compared to birds and mammals (Sotero-Caio et al. 2017).

Short interspersed nuclear elements (SINEs) are a type of nonautonomous retrotransposon (Elbarbary et al. 2016) and have been identified in diverse eukaryotes including mammals, reptiles, fish, ascidians, insects, and flowering plants (Kramerov and Vassetzky 2005). Insertions of SINEs within the new locations around genome may be deleterious to the host (Kazazian, 2004), though the small size of SINEs compared to their long interspersed nuclear elements counterpart, an autonomous retrotransposon, may result in a greater tolerance of their presence (Elbarbary et al. 2016). Often occurring in excess of 10^5 copies/cell (Hamada et al. 1997), the mobilization and recombination of SINEs within the genome have resulted in interspecific diversity and intraspecific polymorphisms via insertions (Kazazian 2004); further, copies are not all identical, and their sequences can differ by 5 – 35% (Kramerov and Vassetzky 2011). Due to their widespread distribution in diverse eukaryotes, low risk of host deleterious effects compared to other retrotransposons, high copy number per cell, and significant sequence variation, SINEs emerge as promising DNA markers to provide valuable insights into interspecific diversity and intraspecific polymorphisms.

Endangered Atlantic Whitefish (*Coregonus huntsmani*) was used as a test case for the development of SINE-based eDNA tools due to the potential of this novel marker type to confer an increase in sensitivity relative to commonly used mtDNA markers for species detection (Chapter 2). *Sma*I-cor is a SINE present within the Coregoninae subfamily and is closely related to the *Sma*I SINE found in Pink Salmon (*Oncorhynchus gorbuscha*) and Chum Salmon (*Oncorhynchus keta*) (Hamada et al. 1997). *Sma*I-cor contains two subtypes, I and II, defined by their differences to *Sma*I. Hamada et al. (1997) found inter- and intra-specific variation in *Sma*I-corI and *Sma*I-corII, however Atlantic Whitefish was not tested.

While *Sma*I-corII provided greater detection sensitivity than mtDNA in eDNA applications (Chapter 2), the qPCR analyses used did not provide insights to the amount of sequence variation within this SINE in Atlantic Whitefish. Here, the sequence variation of *Sma*I-corII in Atlantic Whitefish was explored. Identification of amplicon sequence variants (ASVs) requires separation of true biological variation from PCR or sequencing errors (Couton et al. 2021). Though many pipelines exist to analyze eDNA data, the majority have been developed to process metabarcoding data with the goal of detecting organisms at the species level (Mathon et al. 2021; see Appendix E for a detailed description of eDNA metabarcoding pipeline processes). Here, sequencing methods have been employed to search for intraspecific genetic variation in eDNA samples, and as such pipeline validity for SINE analysis needed to be assessed. This chapter firstly validated pipeline suitability for detection of *Sma*I-corII ASVs. Secondly, ASVs were assessed at two hierarchical levels by comparing *Sma*I-corII variants within Atlantic Whitefish and between Atlantic Whitefish and Lake Whitefish (*Coregonus clupeaformis*) samples to determine (i) if *Sma*I-corII variants can be used to detect genetic variation within Atlantic Whitefish and (ii) if the species-specificity of this marker can be increased by identifying variants unique to either species.

3.3 <u>Methods</u>

3.3.1 Sequence library preparation

To identify ASVs of the *Sma*I-corII region, previously extracted genomic (g)DNA samples from 16 wild caught Atlantic Whitefish housed at Dalhousie's Aquatron were assessed along with eDNA samples from the Aquatron (2.3.1) and 0 m net pen release samples (2.3.4). To assess *Sma*I-corII ASV differences between Atlantic Whitefish and Lake Whitefish to identify species-specific variants, aqueous eDNA samples were collected from Shingle Lake, NS, where Lake Whitefish occur (Hasselman et al. 2009). Sample collection and processing followed the procedure described in section 2.3.4.

DNA libraries were prepared for each sample type by first amplifying the extracts in triplicate PCR replicates in a 10 µL PCR in which each reaction contained 5 µL 2X Invitrogen™ Platinum™ SuperFi II Master Mix (Thermo Fisher Scientific), 0.5 µL 10 µM *Sma*I-corII primers, 3.5 µL water, and 1 µL gDNA. PCR cycling parameters were 98°C 2 min, 40x (98°C 15 s, 60°C 30 s, 72°C 15 s), 72°C 5 min, 10°C hold.

Following amplification, samples underwent an indexing PCR to add unique DNA indices to each sample within the library. Indexing used the Phusion ® High-Fidelity DNA polymerase kit (#M0530L, New England BioLabs). Each indexing PCR comprised a 10 µL final volume that included 2 µL 5X Phusion ® HF Buffer, 1 µL 2mM dNTPs, 0.5 µL PCR product, 0.5 µL xGen™ Normalase™ UDI Primers (#10009797, Integrated DNA Technologies), and 5.9 µL molecular grade water. Index PCR cycling parameters were 98°C 2 min, 20x (98°C 10 s, 62°C 30 s, 72°C 15 s), 72°C 10 min, 10°C hold. Following indexing, each library was pooled in a single tube.

Libraries were bead cleaned to remove small DNA fragments following indexing. Briefly, 50 µL room temperature AMPure XP magnetic beads (#A63882, Beckman Coulter) and 50 µL DNA library were added to one well of a strip tube. After 15 min, the tube was placed on a magnet and left for 5 min. The liquid was removed from the tube before tube contents were washed twice with 200 µL fresh 80% ethanol. The tube was dried at 37°C for 3 min before 50 µL 10mM-Tris/0.05%-Tween was added. After 2 min on the bench, the tube was placed on the magnet for a further 5 min. The liquid was then moved to a new tube and the library was stored at -20°C.

Prior to sequencing, libraries were quantified with the KAPA Library Quantification Kits (#KK4953, Roche). Clean libraries were diluted 10,000x using 10mM-Tris/0.05%-Tween. Each well of the qPCR contained 4.5 µL KAPA Sybr Green Master Mix, 2 µL library, and 1 µL molecular grade water. Cycling parameters were 95□C 5 min, 35x (95□C 30 s, 60□C 45 s), then hold at 37□C. If library concentrations were > 4 nM sequencing proceeded. Libraries which did not meet the concentration requirement were remade. Owing to the higher quality of the gDNA samples compared to eDNA, the gDNA library underwent single-end sequencing to maximize the number of samples which could be sequenced on the run. eDNA libraries underwent paired-end sequencing to ensure the low-quality sequences at ends of reads would be removed during bioinformatic processing. All sequencing used Illumina v2 300-cycle kits on an Illumina MiSeq System.

3.3.2 Bioinformatic pipeline optimization and comparison for *Sma*I-corII ASV analysis

To date, most eDNA metabarcoding pipelines have been limited to species detection by identifying differences within a highly conserved region to assign sequences to species rather than interpret intraspecific variation (Andres et al. 2023a). To determine ASVs of the *Sma*I-corII marker, it was imperative to assess how different pipelines would analyze *Sma*I-corII sequences, a variable SINE. Two commonly used pipelines were chosen, a custom UNIX-based pipeline (hereby called “custom”) and DADA2 (Callahan et al. 2016). The custom pipeline algorithm followed the processes described in Appendix E, while the DADA2 pipeline differed in some regards. Notably, the DADA2 algorithm uses a parametric error model to estimate the Illumina sequencing error rate of the dataset (Callahan et al. 2016). Like the denoising process described in Appendix E, during denoising DADA2 creates centroids based on sequence abundance and iteratively partitions sequences either under the centroid or into a new cluster. DADA2 differs from the custom pipeline as this partitioning uses the estimated error rate calculated earlier in the pipeline to determine the probability that a given sequence is an error of the centroid based on its abundance (Callahan et al. 2016). The OMEGA_C parameter corresponds to the threshold required to correct inferred erroneous sequences within the cluster to match the centroid sequence.

Pipeline suitability for *Sma*I-corII analysis was determined by assessing the number of ASVs detected and total read depth retained. The pipelines were first run using default parameters on the Atlantic Whitefish gDNA samples which were sequenced in triplicate PCR replicates. As gDNA is generally high in quality and concentration, it was assumed true ASVs would be consistently detected across the PCR replicates of an individual fish whereas false positives due to PCR or sequencing error would occur stochastically.

The custom pipeline followed the Bioinformatic Methods for Biodiversity Metabarcoding tutorial (Creedy, Vogler, and Penlington, https://learnmetabarcoding.github.io/LearnMetabarcoding) and was first assessed with default parameters (Table 6). Forward and reverse primers were removed from demultiplexed sequences with CUTADAPT v.3.5 which allows 10% of base mismatches between the query and primer sequences by default (Martin 2011). Though the priming region was expected to be at the beginning of the read, primers were not anchored to account for the possibility that they were further into the sequence. Sequence quality was assessed with FASTQC v.0.11.9 (Andrews

2010). As the *Sma*I-corII region was expected to be variable (Hamada et al. 1997) but the details of variation within Atlantic Whitefish were unknown, sequences were not trimmed to a fixed length to account for length variations due to insertions and deletions (indels). Sequences were concatenated into a single fastq file and quality filtered using the --fastx_filter command in VSEARCH and the --fastq_maxee default value of 1 (Rognes et al. 2016). The concatenated fastq file was then dereplicated using the --derep_fulllength command and the --minlen 32 parameter, which discards sequences shorter than 32 bases long. Dereplicated sequences were then denoised using the UNOISE3 algorithm (Edgar 2016a) implemented in VSEARCH with parameters --minsize 8 --unoise_alpha 2. Chimeric sequences were removed using the UCHIME3 (Edgar 2016b) algorithm implemented in VSEARCH using the --uchime3_denovo command. Remaining reads were mapped to ASVs with the VSEARCH --search_exact command. Taxonomy was assigned to the ASVs against a custom database containing the *Sma*I-corII consensus sequence (Hamada et al. 1997) using the blastn function of BLAST (Madden 2002) with the parameters -num_threads 1 -evalue 0.001 -perc_identity 97.

Table 6 Custom pipeline parameter descriptions and default values.

Pipeline step and program	Parameter	Definition	Default value
Primer removal: CUTADAPT	e	Maximum error rate (percentage) in priming region	0.1
Quality filtering: VSEARCH	--fastq_maxEE	Expected error rate (incorrect base call). Sequences with error rates higher than this value will be discarded	1
Dereplication: VSEARCH	minlen	Sequences shorter than this value will be discarded	32
Denoising: VSEARCH	minsize	Sequences which occur less than this value will be discarded	8
	unoise_alpha	Dissimilarly level during clustering	2
Assigning taxonomy: BLAST	perc_identity	Percentage of query sequence matching the reference sequence	97

The DADA2 pipeline operates within the R environment and was executed in RStudio v4.2.0. As with the custom pipeline, DADA2 was first run with default parameters (Table 7), following the DADA2 ITS tutorial (https://benjjneb.github.io/dada2/ITS_workflow.html). Ambiguous bases (Ns) were removed using the filterAndTrim command with parameter maxN = 0. Forward and the reverse-compliment of the reverse primer were removed from sequences with CUTADAPT v3.5. Read quality profiles were plotted to assess where along the sequence base pair (bp) quality dropped off. To account for the unknown length variation of the *Sma*I-corII marker, reads were not trimmed to a consistent length. Quality filtering used the filterAndTrim command with parameters maxN = 0, maxEE = 2, truncQ = 2, minLen = 50. The error rate of the dataset was then estimated from a subset of samples with the learnErrors function. Samples were dereplicated with the derepFastq function. Dereplicated samples then underwent DADA2's core sample inference algorithm (denoising) which included the previously learned error rate as well as parameters OMEGA_A = 1e-40, OMEGA_C = 1e-40, BAND_SIZE = 16. An ASV table was then created from the denoised samples using the makeSequenceTable function. Chimeras were removed from the samples with the removeBimeraDenovo function, using method = consensus. Taxonomy was assigned to the ASVs against a custom database containing the *Sma*I-corII consensus sequence using the assignTaxonomy function.

Table 7 DADA2 parameter descriptions and default values.

Pipeline step and program	Parameter	Definition	Default value
Primer removal: CUTADAPT	e	Maximum error rate (percentage) in priming region	Unspecified but assumed to be CUTADAPT default of 0.1
Quality filtering: DADA2	maxN	Maximum number of ambiguous bases allowed. Sequences with error rates higher than this value will be discarded	0
	maxEE	Expected error rate. Sequences with error rates higher than this value will be discarded	2
	truncQ	At the first instance of a quality score less than this value reads will be truncated	2
	minLen	Sequences shorter than this value will be discarded	50
Denoising: DADA2	OMEGA_A	Threshold for calling a unique sequence significantly abundant (to make a centroid)	1e-40
	OMEGA_C	Threshold for sequences inferred to contain errors to be corrected in the final output	1e-40
	BAND_SIZE	Restricts the total number of indels within a sequence relative to another	16
Assigning taxonomy	assignTaxonomy	-	-

After each pipeline was run, the total read depth retained after each step, the number of ASVs, and if the ASV was consistently detected across the PCR replicates of a single fish was assessed and compared. The pipelines were then iteratively run, changing one parameter at a time to assess how the suite of ASVs and total read depth per ASV changed in response (Figures F1 and F2; see Appendix F).

Following the initial parameter analysis, errors within the sequence files which skewed the optimization were identified. Notably, the sequencing run included duplicated indices which resulted in the concatenation of Atlantic Whitefish and *Lobelia* spp. sequences within the same demultiplexed file. During initial optimization, the custom pipeline showed a large drop in read depth compared to DADA2 following the primer removal (see Figure 5), and optimization was concentrated on increasing the number of reads retained during this step. Nonetheless, the identification of *Lobelia* spp. sequences within the demultiplexed files highlighted a key difference in how the two pipelines treat erroneous sequences as the DADA2 pipeline gave no indication of the error. DADA2 was then run with the error correction parameter (OMEGA_C) turned off and the number of reads retained dropped by 70%, indicating that by correcting erroneous sequences to match the centroid, the depth of the centroid was being artificially inflated. Optimization was then restarted for the custom pipeline only, with the understanding that the initial read count within the samples was reflective of both Atlantic Whitefish and *Lobelia* spp. sequences.

3.3.3 Custom bioinformatic pipeline optimization and *Sma*I-corII ASV determination

Using default parameters, DADA2 identified higher read depths across fewer ASVs than the custom pipeline. When the OMEGA_C error correction parameter was turned off in the DADA2 pipeline, the two pipelines retained a similar number of ASVs and similar depths following denoising (Appendix F). The error estimation and correction functions of the DADA2 pipeline were therefore found to be unsuitable for analysis of *Sma*I-corII and the custom pipeline was chosen for further optimization. Custom pipeline outputs (depth, number of ASVs, and ASVs consistency) were again assessed firstly using default values of all parameters, where ASV consistency was used as a metric of how optimized the pipeline was. Analysis of the output resulted in addition of a quality trimming step prior to concatenation and dereplication, where a minimum phred score of 26 and minimum and maximum sequence length limits (-q 26 -m 70 -M 105) were enforced based on sequence lengths identified in the initial discovery of *Sma*I-corII (Hamada et al. 1997). Sequences were trimmed at the first instance of a base with a phred score lower than 26 (1 in 400 chance of base call error) and any sequences outside the specified length range were dropped. Following quality trimming, cluster contents from the custom pipeline were assessed during the denoising stage by adding the function --uc which produced a tab-separated output file of centroids and sequences within the cluster.

Analysis of the custom pipeline cluster content following denoising revealed ASVs differing from the centroid by only one or a few single base changes (SNPs) were being clustered, thereby hiding them in the final pipeline output. As the goal of the optimization was to identify such variants, the custom pipeline was therefore stopped after the dereplication step to ensure these variants were not lost. The post-dereplication output of the custom pipeline consisted of a table of over 275,000 unique ASVs and their corresponding occurrence counts per sequenced PCR replicate. Downstream analysis was undertaken in RStudio v.4.2.0 in lieu of denoising to determine which ASVs represented true biological variation, and which were residual PCR/sequencing errors.

For ASVs determined from the gDNA samples used during pipeline optimization, different depth cut-off thresholds were assessed to remove low quality PCR replicates prior to any ASV analysis. Samples were inferred to be low quality if the total depth of a PCR replicate was vastly lower than the other PCR replicates originating from the same individual fish (Figure 4A). Thresholds of 10%, 1%, and 0.1% total depth were applied on a per PCR replicate basis to the entire sample set, whereby samples needed to have a depth equal to or greater than the threshold to be retained. Owing to the generally higher quality of gDNA compared to eDNA, ASVs were then assessed across the triplicate PCR replicates of an individual fish under the assumption true ASVs would be consistently detected across PCR replicates whereas false positives due to PCR or sequencing error would occur stochastically. Errors were also assumed to occur at lower depths than true ASVs. Therefore, only ASVs that had depths of at least 1% of each PCR replicate of a single fish were retained. ASVs which passed the two downstream filtering thresholds (PCR replicate depth cut-off and ASV minimum depth) were considered to be putatively true ASVs.

Following determination of optimal downstream filtering metrics, the custom pipeline optimization was re-evaluated to determine if streamlining pipeline parameters would reduce the number of putatively false ASVs present in the dereplicated output. As many singleton ASVs were present in the dereplicated output, the --minuniquesize 10 parameter was added during dereplication, where ASVs needed to have a depth of at least 10 to be retained in the final dereplicated output (Jensen et al. 2020). Increasing the number of allowable mismatches in the priming region from the default 10% (e 0.1) to 20% (e 0.2) was assessed as was increasing --fastq_maxee 1 (default) to 2 to determine if less stringent primer matching and quality filtering, respectively, increased the number of ASVs detected in the final output following downstream filtering. A 2-sample test for equality of proportions with continuity correction was conducted to compare the proportion of reads retained between different parameter values.

The optimal parameters values determined via gDNA assessment were applied to the ASV analysis of eDNA samples to determine if differences in the suite of ASVs detected would vary between sample types. As with the gDNA samples, a 1% depth cut-off threshold was firstly applied to the eDNA samples on a per PCR replicate basis to remove low quality PCR replicates. Due to the heterogeneous distribution of aqueous eDNA, where the DNA of multiple individuals can co-occur within the water, putatively true ASVs could not be determined by assessing consistency across PCR replicates as ASV detections could not be assigned to any one fish (Figure 4B). Following the first depth cut-off, a second 1% depth cut-off threshold was applied on a per ASV basis, where the total depth of a given ASV across all remaining PCR replicates needed to be greater than or equal to 1% of the total remaining depth.

Following optimization of the custom pipeline and downstream ASV determination methods, the ASV output from these optimized protocols were again compared to DADA2, which was run with the addition of a maximum sequence length enforcement of 105 bp during the trimming step and without its error correction function during denoising to be consistent with the custom pipeline parameterization and to confirm pipeline suitability. Putatively true ASVs determined by the custom pipeline and confirmed by the comparison to the DADA2 output were added to a custom reference database alongside the *Sma*I-corII consensus sequence (Hamada et al. 1997). Following pipeline confirmation and ASV determination within the gDNA samples, Aquatron eDNA and 0 m net pen eDNA samples were assessed to determine if ASVs differed between Atlantic Whitefish sample types. The eDNA samples were bioinformatically assessed with the optimized custom pipeline the same way as the gDNA samples with one modification: as the eDNA samples underwent paired-end sequencing (3.4.1) forward and reverse reads were merged with PEAR v0.9.11 using a quality threshold parameter of -p 26 and minimum overlap of -v 18 which occurred after sequence quality checking with FASTQC and before length trimming with CUTADAPT. eDNA samples from Shingle Lake were also assessed for *Sma*I-corII ASV variation to determine if any putatively true ASVs were unique to the samples containing Atlantic Whitefish DNA (gDNA, Aquatron eDNA, and 0 m net pen eDNA) or Lake Whitefish DNA (Shingle Lake eDNA), therefore bringing the specificity of the marker down from subfamily-level to species-level. Welch's t-tests were performed in RStudio v.4.2.0 to compare the proportional depth of the ASVs detected within both species. ASVs were aligned with MUSCLE (Edgar 2004).

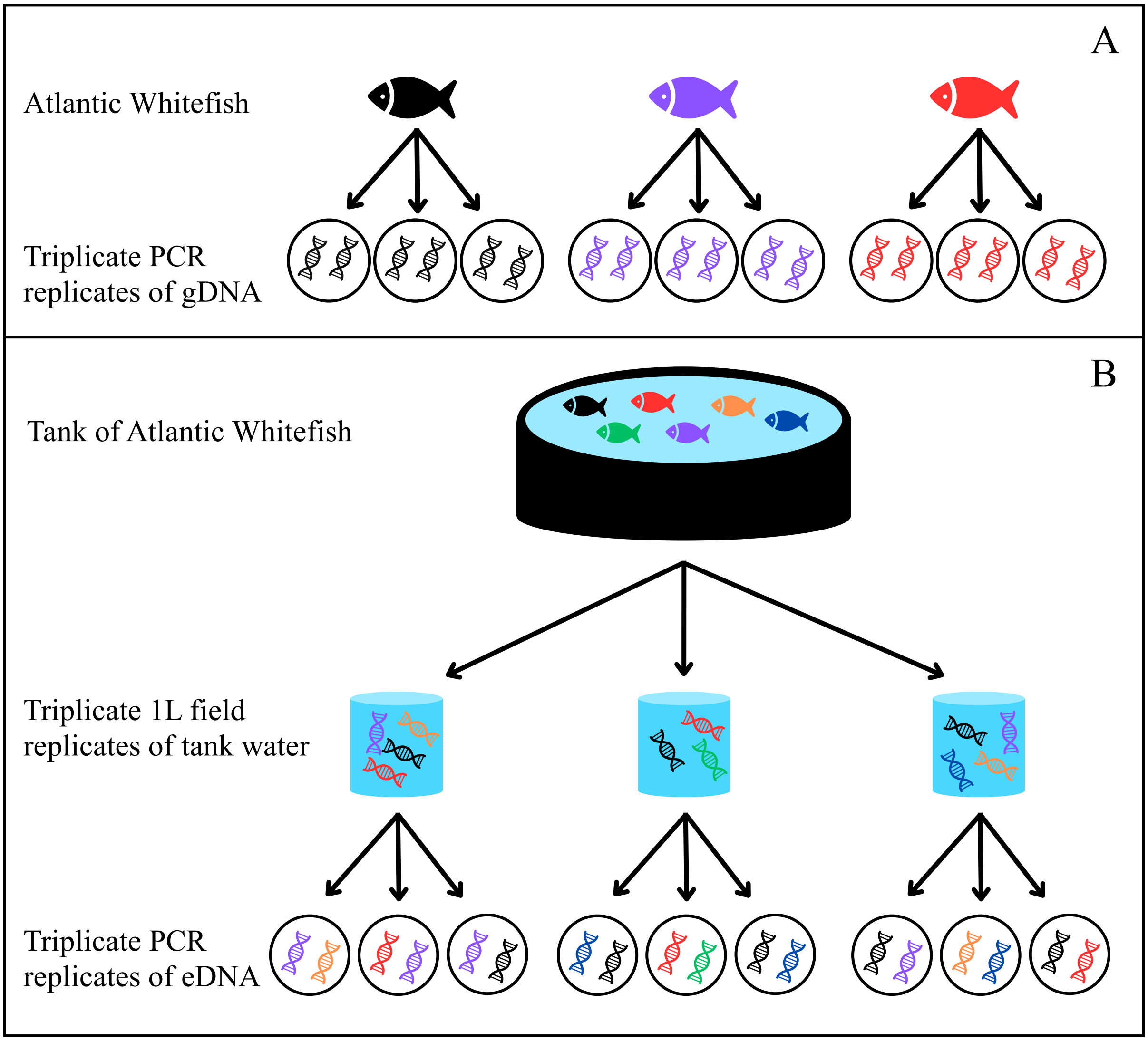
A
Atlantic Whitefish
Triplicate PCR replicates of gDNA
B
Tank of Atlantic Whitefish
Triplicate 1L field replicates of tank water
Triplicate PCR replicates of eDNA

Figure 4 Comparison of gDNA (A) and eDNA (B) sample collection and rationale for different processing methods for SmaI-corII ASV determination.

3.3.4 Tank dilution series

Following the creation of a custom reference database containing putatively true *Sma*I-corII ASVs, a dilution series was performed to assess variant drop out at different eDNA concentrations. Total DNA of the Aquatron eDNA extracts (2.3.1) was quantified with a NanoDrop before extracts underwent five rounds of 10-fold dilutions with molecular grade water. A sequencing library was created from the dilutions following the procedure described in section 3.3.1.

Stochastic effects during PCR increase with decreasing DNA concentrations (Weusten and Herbergs 2012) and low concentrations can also lead to false mutations (Akbari et al. 2005). Therefore, multiple analyses performed in RStudio v.4.2.0 were compared to determine an optimal method to determine the ASV composition of the dilution series samples following bioinformatic analysis with the optimized custom pipeline (3.3.3) (Table 8). Following the analyses, the number of ASVs and total depth per analysis output were compared to determine the most suitable method for ASV assessment across decreasing concentrations, where high depth across the ASVs previously identified (3.3.3) were considered markers of analysis suitability.

Table 8 Methods tested and compared to determine ASV composition of the tank dilution series where optimized custom pipeline refers to stopping after the dereplicate stage while custom pipeline refers to running the pipeline in its entirety.

Analysis attempt	**Analysis description**
1	Optimized custom pipeline (2.3.2) and downstream eDNA analysis (3.3.3)
2	Optimized custom pipeline (2.3.2) and only retained ASVs which were a 100% match to the custom reference database of *Sma*I-corII ASVs
3	Optimized custom pipeline, used the custom reference database of *Sma*I-corII ASVs when assigning taxonomy with BLAST (-perc_identity = 100)
4	Same as analysis 3, however only retained ASVs with 100% query coverage

3.4 Results

3.4.1 Bioinformatic pipeline optimization and comparison for *Sma*I-corII ASV analysis

Sequencing of Atlantic Whitefish gDNA extracts returned 2,197,228 single-end reads. For the custom pipeline run with default settings, 1,261,151 reads remained following primer removal, 1,230,551 after quality filtering, 1,206,831 after dereplication, 241,377 after denoising, and 240,716 after chimera removal. When assigning taxonomy with BLAST by comparing the reads to the *Sma*I-corII consensus sequence (Hamada et al. 1997), 166,197 reads were retained (Figure 5). When DADA2 was run with default settings, all 2,197,228 reads remained following priming removal as CUTADAPT does not discard untrimmed reads unless specified. The number of reads remaining was 1,890,741 after quality filtering, 1,873,342 after denoising, and 1,868,025 after removing chimeric sequences, all of which were retained following taxonomy assignment.

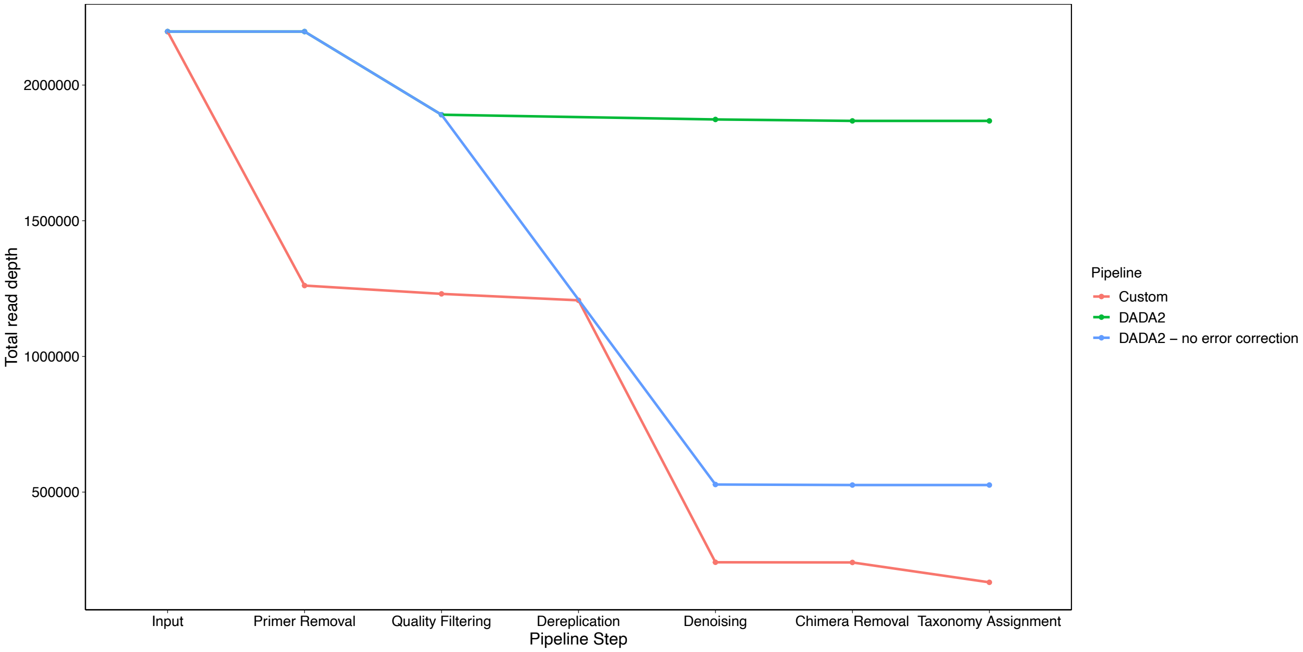

Figure 5 Read tracking of the custom and DADA2 pipelines using default settings and DADA2 without its error correction parameter to assess read depth of ASVs across Atlantic Whitefish gDNA sequences.

In assessing the consistency of ASVs detected by each pipeline, the custom pipeline run with default settings had a higher ratio of consistently detected ASVs across PCR replicates of an individual fish (211/373) than DADA2 (14/1594); however, the total read depth of the consistently detected ASVs with the custom pipeline was 86% lower than the read depth retained by DADA2 at 159,047 reads compared to 1,165,763 reads, respectively. Of these ASVs, the majority were 151 bp in length, much longer than the 90 bp expected amplicon length of the *Sma*I-corII assay (Figure 6).

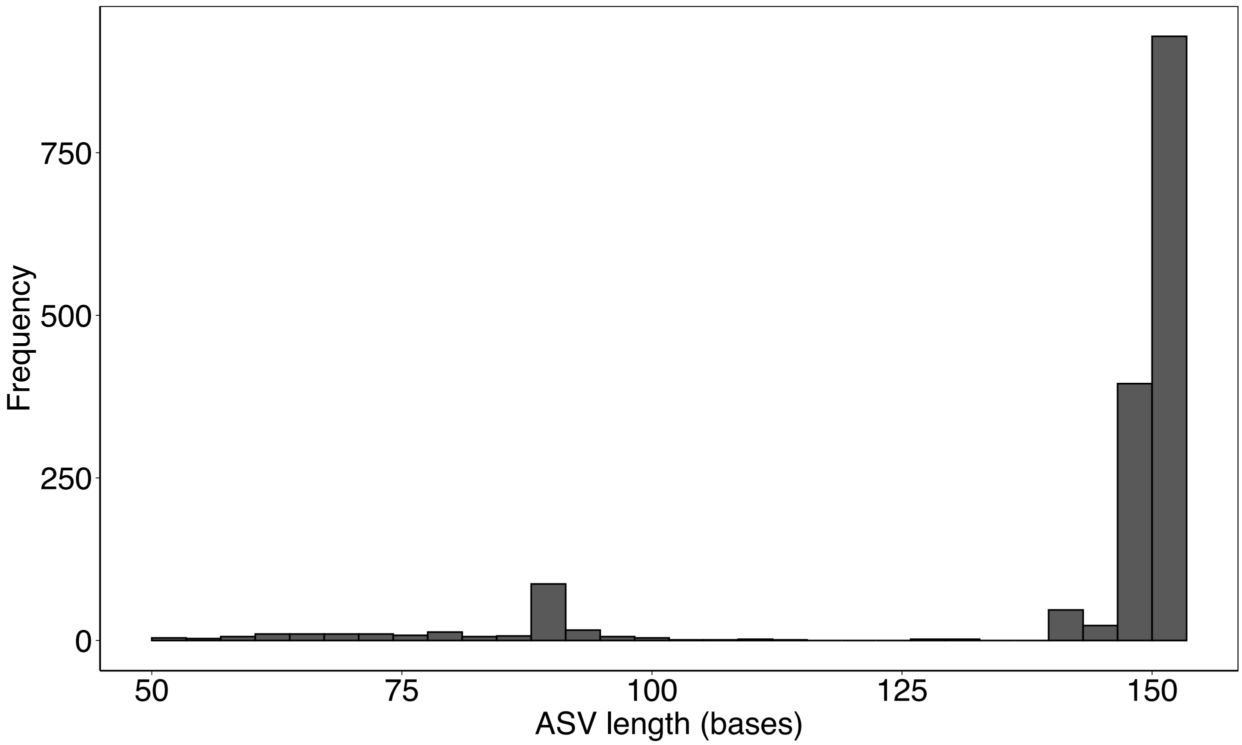

Figure 6 ASV length distribution of the DADA2 pipeline run with default parameters.

The large difference in total read depth, number of ASVs detected, and ratio of consistency across ASVs indicated that running either pipeline with default parameters was not optimal for a marker as variable as a SINE. The large difference in retained reads between the pipelines can be attributed to DADA2's error correction parameter (OMEGA_C), which corrects inferred erroneous sequences to match their centroid sequence. When this parameter was turned off, the number of retained reads dropped to 526,049 reads 1,598 ASVs in the final DADA2 output (see Appendix F). Of these, 14 ASVs were consistently detected across PCR replicates (0.88%) totaling 168,389 reads (32%). As ASV consistency was considered a proxy for pipeline suitability, DADA2 was therefore determined to be unsuited for *Sma*I-corII variant assessment

As a large proportion of reads were lost with the custom pipeline during taxonomy assignment via BLASTing against the *Sma*I-corII reference sequence (31% following chimera removal), ASVs prior to this step were assessed to determine the necessity of this method of taxonomy assignment for a targeted marker. Prior to taxonomy assignment, the custom pipeline retained 240,716 reads across 2,900 ASVs. Of these, 1,574 ASVs (54%) comprising 188,419 reads (78%) were consistently detected across PCR replicates. Due to the higher read depth (1.4x) across more ASVs (7.8x) prior to taxonomy assignment, the custom pipeline was only run through the chimera removal step prior to ASV assessment during initial pipeline optimization and comparison to DADA2 as BLAST was found to be unsuited for *Sma*I-corII variant assessment (see Appendix F for initial pipeline optimization parameters and results).

3.4.2 Custom bioinformatic pipeline optimization and *Sma*I-corII ASV determination

The custom pipeline output with default parameters included a high proportion of ASVs shorter than the 90 bp expected amplicon length of the *Sma*I-corII assay (Figure 7). Analysis of sequence quality via the FASTQC report indicated the per base phred scores dropped below 26 around 105 bp into the sequence length. The distribution of sequence lengths and sequence quality drop off resulted in the addition of a read trimming step in the pipeline. Enforcing length limits of 70 – 105 bp to account for indels resulted in a final pipeline output of 210,153 reads across 2495 ASVs following chimera removal, the majority of which were the expected amplicon length of 90 bp.

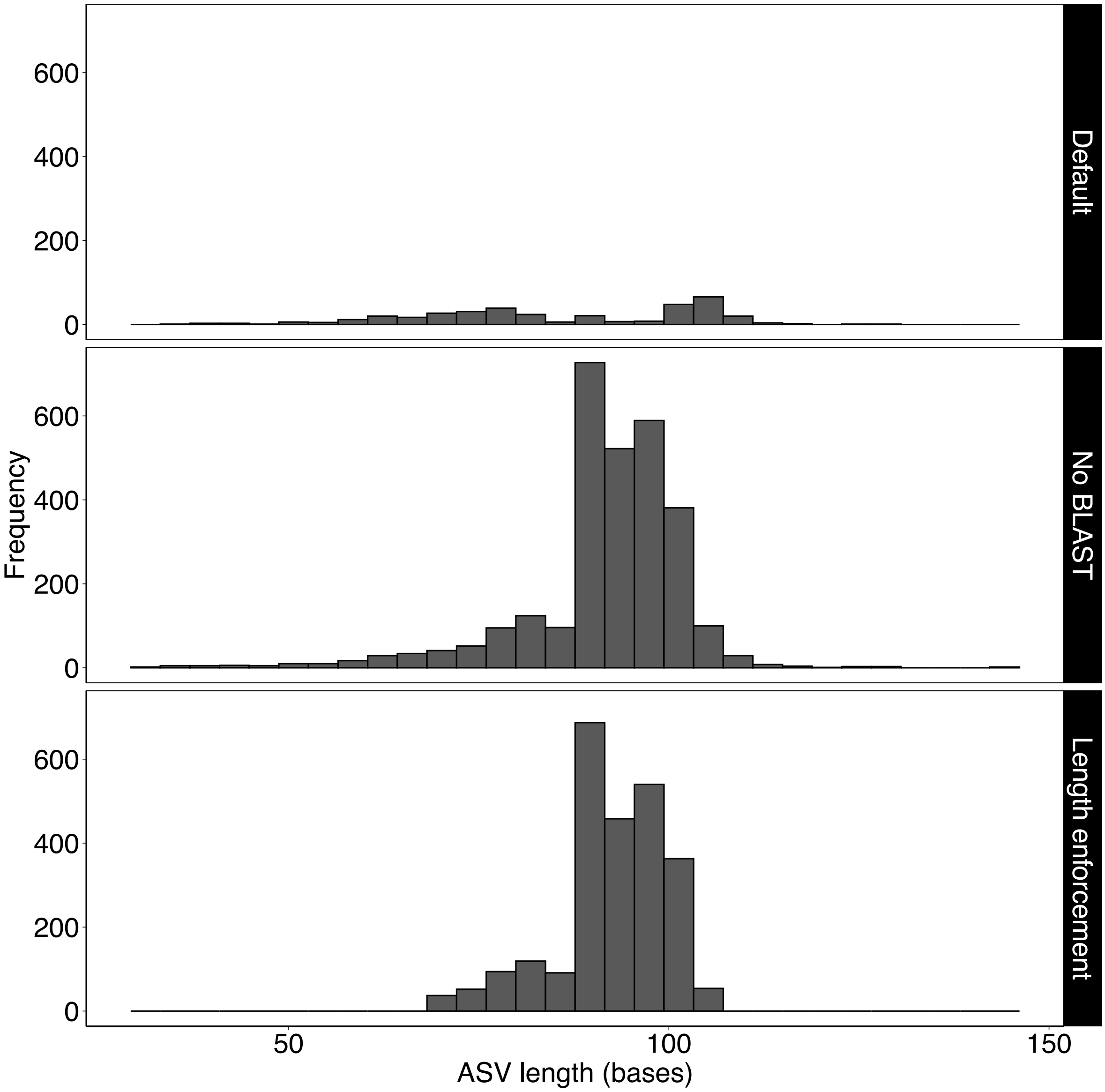

Figure 7 ASV length distribution of the custom pipeline output run using default parameters through taxonomy assignment via BLAST compared to both omitting taxonomy assignment via BLAST as well as adding length enforcements during quality trimming and omitting taxonomy assignment via BLAST.

For the default custom pipeline, the largest drop in read depth was observed following the denoising step (Figure 5). This step involves the application of a sequence clustering and comparison method to identify and eliminate residual PCR/sequencing errors. The method evaluates the number of base differences and abundance ratios among sequences, effectively inferring and subsequently removing errors (see Appendix E for a more detailed description of clustering). Analysis of cluster content following length enforcement revealed 10,704 sequences with depths of 8 – 13,106 reads were clustered within the *Sma*I-corII consensus sequence (Hamada et al. 1997) centroid, the most abundant ASV at 121,305 reads. The high depths of some of these sequences (10.83% of the consensus sequence depth) indicated denoising was not an effective method of *Sma*I-corII ASV determination. Rather, the pipeline was stopped after the dereplication stage, which outputted all unique sequences and their corresponding abundances, and manual downstream depth filters were applied to parse true *Sma*I-corII biological variation from residual PCR/sequencing errors.

The post-dereplication output of the custom pipeline following sequence length enforcement consisted of 275,128 sequences totalling 1,187,971 reads and downstream analysis was undertaken in RStudio v4.2.0. in lieu of denoising to determine putative true ASVs from PCR/sequencing errors. To account for the possibility of low-quality DNA or failed amplification of PCR replicates which could impact ASV determination, a depth-based threshold was firstly applied on a per PCR replicate basis. When a 10% depth threshold was applied across the entire sample set, whereby the depth of a PCR replicate needed to contain 10% of the total depth of the run to be retained, all PCR replicates were removed. The 1% threshold resulted in 3/3 PCR replicates from individual 22-H1 and 1/3 PCR replicates from individual 18-A1 being removed while the 0.1% threshold resulted in 1/3 PCR replicates being removed from each 22-

H1 and 18-A1. For each fish, the ASV corresponding to the *Sma*I-corII consensus sequence (Hamada et al. 1997) had the highest depth of all retained ASVs; as the depths of the remaining PCR replicates from individual 22-H1 were 2.3 and 4.6 standard deviations away from the mean consensus sequence depth following dereplication, the 0.1% threshold was determined to be too lenient, and the 1% threshold was selected. Retaining only sequences which had 1% total depth in all retained PCR replicates of an individual fish resulted in three ASVs totalling 152,017 reads being retained across all individuals.

ASV determination was an iterative process and the number of sequences retained in the dereplicated output strongly influenced the downstream depth-based thresholds. Therefore, following determination of optimal downstream depth-based filtering methods (1% depth cut-off per PCR replicate of the sample set and 1% depth cut-off of consistently detected ASVs per PCR replicate of an individual fish), earlier pipeline parameters related to primer removal, quality filtering stringency, and dereplication were re-evaluated to assess their influence on the dereplicated output. A minimum sequence depth of 10 reads was added during the dereplication stage to remove rare, low abundance sequences which could not reliably be distinguished from PCR/sequencing errors (Jensen et al. 2020). The addition of this minimum sequence occurrence requirement reduced the number of sequences by 95% (13,515 sequences retained) and the total depth by 30% (839,546 reads retained), indicating there were many singletons and rare sequences present in the initial dereplicated dataset which impacted the downstream depth-based filtering. Downstream analysis following the addition of the minimum sequence occurrence resulted the identification of two additional ASVs for a total of five ASVs totalling 172,898 reads.

Increasing the proportion of mismatches allowed in the priming region from 0.1 to 0.2 did not change the number of ASVs detected following downstream depth filtering; however, the read depth of the six ASVs decreased by 3%. The analysis resulted in a Chi-Square statistic (χ^2) of 3.59, with 1 degree of freedom. The associated p-value of 0.058 indicated the difference in proportion of reads retained across the five ASVs did not significantly differ when increasing the error allowance in the priming region from 0.1 to 0.2 and to maintain a conservative error allowance, the default value of 0.1 was selected as optimal for the custom pipeline.

Likewise, increasing the stringency of the –fastq_maxEE quality filtering parameter which refers to the expected error rate of a sequence from the default 1 to 2 resulted in the same five ASVs detected following downstream depth filtering and the depths of the ASVs increased by 3%. The analysis resulted in a Chi-Square statistic (χ^2) of 0.222, with 1 degree of freedom. The associated p-value of 0.638 indicated the difference in proportion of reads retained across the five ASVs did not significantly differ when increasing expected error rate from 1 to 2 and to maintain a conservative approach, the default value of 1 was selected as optimal for the custom pipeline.

To further confirm the suitability of the optimized pipeline and downstream filtering methods for *Sma*I-corII variant detection, a final comparison was made to DADA2. The optimized custom pipeline was ultimately determined to be better suited for *Sma*I-corII variant determination and the five ASVs retained following analysis with the custom pipeline and optimized downstream depth-based filtering were considered to be putatively true (see Appendix F).

When comparing the five ASVs determined with the custom pipeline and optimized downstream filtering thresholds to the custom pipeline run through the denoising step, only

ASVs 01 and 02 formed centroids in the denoised output while ASV 03 – 05 were clustered beneath ASV 01 and hidden in the final ASV list. ASVs 01 – 04 were ubiquitous across all individuals while ASV 05 was not detected in fish 18-A4 and 18-E7 (Figure 8). The relative frequencies (as inferred from sequence depth) of the ubiquitous ASVs were relatively consistent across the individuals. Across the remaining PCR reps from all fish, the total depth retained was 172,898 reads. ASV 01 was the most abundant variant comprising nearly 70% of the total retained read depth across all fish. The remaining depth was split similarly across ASVs 02 (12%), 03 (7.5%), 04 (5.8%), and 05 (5.3%). All five ASVs were observed within eDNA samples from both the Aquatron and 0 m net pen samples at proportions similar the gDNA samples (Table 9). Analysis of the Lake Whitefish eDNA samples from Shingle Lake returned ASVs 01 – 04 as well as two additional ASVs, 06 and 07.

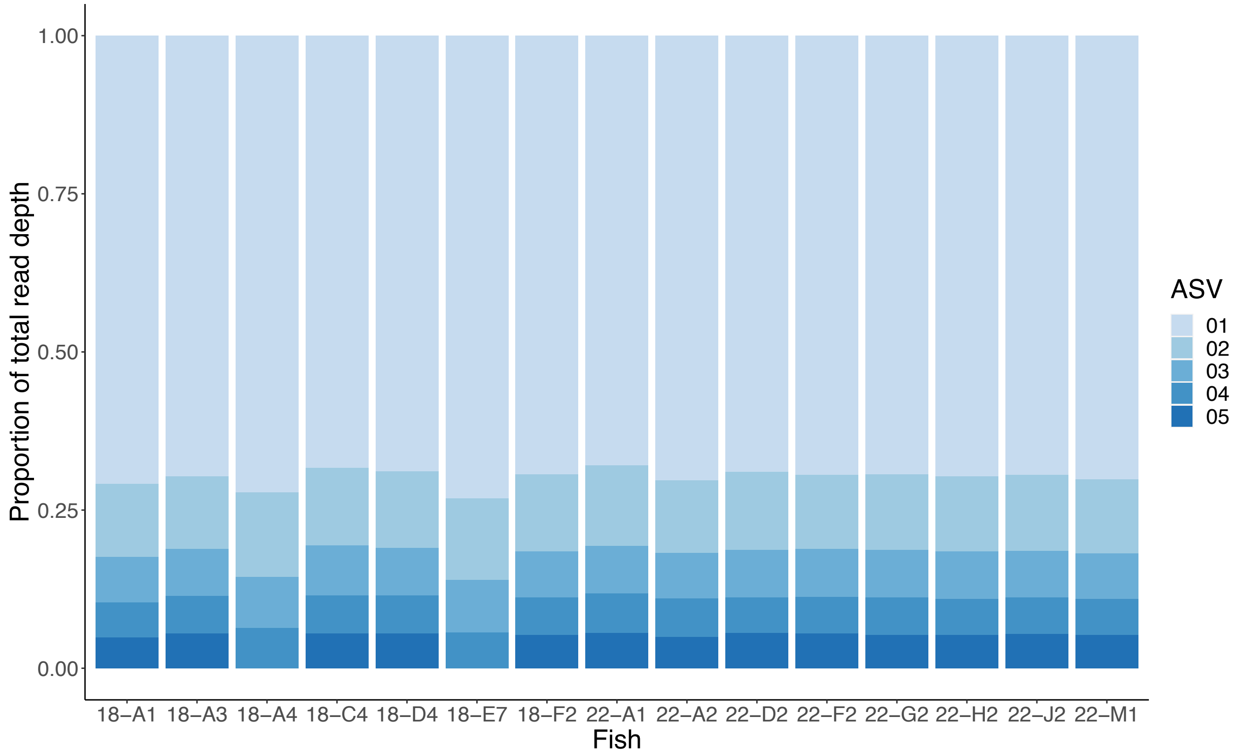

Figure 8 SmaI-corII ASVs detected in gDNA from 15 Atlantic Whitefish following custom pipeline and downstream analysis optimization.

Table 9 Percentage of SmaI-corII ASVs detected in Atlantic Whitefish gDNA or eDNA samples from lakes containing either Atlantic Whitefish or Lake Whitefish.

	Sample	01	02	03	04	05	06	07
Atlantic Whitefish gDNA	18-A1	70.87	11.60	7.13	5.52	4.89	-	-
	18-A3	69.68	11.51	7.39	5.91	5.51	-	-
	18-A4	72.18	13.35	8.13	6.34	-	-	-
	18-C4	68.36	12.27	7.85	6.04	5.48	-	-
	18-D4	68.84	12.16	7.50	6.01	5.49	-	-
	18-E7	73.19	12.85	8.31	5.65	-	-	-
	18-F2	69.37	12.16	7.32	5.90	5.26	-	-
	22-A1	67.91	12.72	7.57	6.19	5.61	-	-
	22-A2	70.30	11.50	7.21	6.03	4.96	-	-
	22-D2	69.00	12.29	7.56	5.58	5.57	-	-
	22-F2	69.46	11.72	7.60	5.70	5.52	-	-
	22-G2	69.35	11.98	7.53	5.86	5.27	-	-
	22-H2	69.69	11.89	7.45	5.72	5.24	-	-
	22-J2	69.45	12.03	7.38	5.70	5.43	-	-
	22-M1	70.20	11.65	7.22	5.69	5.24	-	-
	Mean ± SD	**69.86 ± 1.38**	**12.11 ± 0.53**	**7.54 ± 0.33**	**5.86 ± 0.23**	**5.34 ± 0.23**	-	-
Aquatron eDNA	1	69.26	12.30	7.71	5.86	4.87	-	-
	2	69.68	11.55	8.06	5.90	4.80	-	-
	3	69.69	11.75	7.89	5.86	4.80	-	-
	Mean ± SD	**69.55 ± 0.25**	**11.87 ± 0.39**	**7.89 ± 0.18**	**5.88 ± 0.02**	**4.83 ± 0.04**	-	-
Net pen 0 m eDNA	1	67.60	13.54	6.59	6.12	6.16	-	-
	2	68.66	15.64	5.12	5.60	4.98	-	-
	3	67.51	13.49	7.33	5.96	5.72	-	-
	Mean ± SD	**67.92 ± 0.64**	**14.22 ± 1.23**	**6.35 ± 1.13**	**5.89 ± 0.26**	**5.62 ± 0.60**	-	-
Shingle Lake eDNA*	1	51.47	6.54	9.95	9.23	-	10.04	12.77
	2	58.84	9.25	12.15	11.20	-	6.95	1.61
	3	63.06	7.71	7.60	5.51	-	9.66	6.45
	4	57.17	8.90	9.30	3.92	-	11.01	9.71
	5	60.75	7.50	6.47	7.40	-	10.99	6.90
	Mean ± SD	**58.27 ± 4.38**	**7.98 ± 1.20**	**9.10 ± 2.19**	**7.45 ± 2.89**	-	**9.73 ± 1.66**	**7.49 ± 4.15**

** Shingle Lake samples 1–2 and 3–5 are field replicates of sites 1 and 2, respectively. Field replicate 1 of site 1 failed to amplify.*

Three of the seven detected *Sma*I-corII ASVs appear species-specific, with ASV 05 only detected in Atlantic Whitefish samples and ASVs 06 and 07 only detected in the Lake Whitefish samples (Figure 9). The mean proportional depth of shared ASVs across all Atlantic Whitefish samples were compared to the proportions of ASVs within the Lake Whitefish Shingle Lake eDNA samples. The proportion of ASVs 01 and 02 differed significantly between the two species (ASV 01: t(4.18) = 5.69, p = 0.004; ASV 02: t(5.63) = 8.21, p = 0.000, Welch's t-test) while ASVs 03 and 04 did not (ASV 03: t(4.17) = -1.69, p = 0.164; ASV 04: t(4.01) = -1.23, p = 0.287, Welch's t-test).

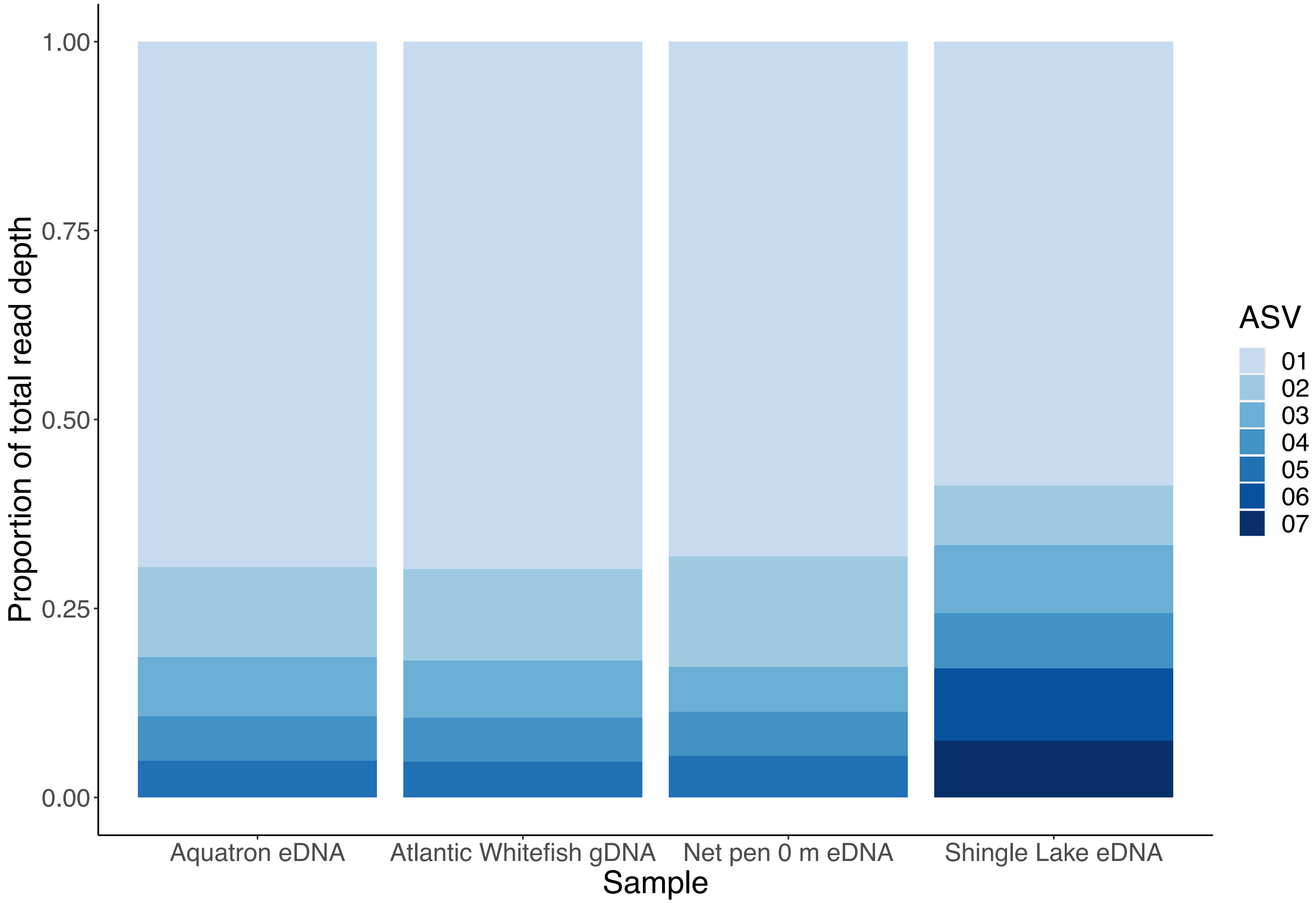

Figure 9 Proportion of *Sma*I-corII ASVs detected in Atlantic Whitefish samples (gDNA, Aquatron eDNA, net pen 0 m eDNA) and Lake Whitefish samples (Shingle Lake eDNA).

ASV 01 was identical to the *Sma*I-corII consensus sequence determined by Hamada et al. (1997) in the initial *Sma*I-cor discovery. ASVs 02, 05, and 06 were differentiated by single base changes: a T → C substitution at location 74, a G → A substitution at location 30, and a C → T substitution at location 3, respectively (Figure 10). ASVs 03, 04, and 07 were defined by insertions of one, two, or three adenine (A) bases at locations 55, 54, and 53 respectively.

```
          1        10        20        30        40        50        60        70        80        90
ASV 01    GGCGCTTGTAACGCCAAGGTAGTGGGTTCGATCCCCGGGACCACCCATACAC---AAAAATGTATGCACGCATGACTGTAAGTCGCTTTGGAT
ASV 02    GGCGCTTGTAACGCCAAGGTAGTGGGTTCGATCCCCGGGACCACCCATACAC---AAAAATGTATGCACGCACGACTGTAAGTCGCTTTGGAT
ASV 03    GGCGCTTGTAACGCCAAGGTAGTGGGTTCGATCCCCGGGACCACCCATACAC--AAAAAATGTATGCACGCATGACTGTAAGTCGCTTTGGAT
ASV 04    GGCGCTTGTAACGCCAAGGTAGTGGGTTCGATCCCCGGGACCACCCATACAC-AAAAAAATGTATGCACGCATGACTGTAAGTCGCTTTGGAT
ASV 05    GGCGCTTGTAACGCCAAGGTAGTGGGTTCAATCCCCGGGACCACCCATACAC---AAAAATGTATGCACGCATGACTGTAAGTCGCTTTGGAT
ASV 06    GGTGCTTGTAACGCCAAGGTAGTGGGTTCGATCCCCGGGACCACCCATACAC---AAAAATGTATGCACGCATGACTGTAAGTCGCTTTGGAT
ASV 07    GGCGCTTGTAACGCCAAGGTAGTGGGTTCGATCCCCGGGACCACCCATACACAAAAAAAATGTATGCACGCATGACTGTAAGTCGCTTTGGAT
```

Figure 10 Alignment of SmaI-corII ASVs determined from Atlantic Whitefish samples (ASVs 01 – 05) and Lake Whitefish samples (ASVs 01 – 04, 06, 07) where ASV 01 represents the *Sma*I-corII consensus sequence (Hamada et al. 1997) and the underlined bases indicate where the hydrolysis probe used in the qPCR assay binds (Chapter 2). Forward and reverse primers flanked the ASVSs are were removed as part of the bioinformatic pipeline for ASV identification.

3.4.3 Tank dilution series

To assess the ability to detect *Sma*I-corII variants at different DNA concentrations, a dilution series was created by serially diluting Aquatron eDNA samples in five 10-fold increments from a starting concentration of 3.02 ng/µL. To account for the possibility that the lowest concentration samples had a greater risk of PCR errors, multiple methods were compared to analyze the amplicons obtained from the dilutions, each of which began by running the dilution series amplicons through the optimized custom pipeline using the parameters described for the eDNA analysis in section 3.3.3. Each sample in the tank dilution series was amplified with *Sma*I-corII and two mitochondrial markers and sequenced in a single run as part of another analysis (Appendix G). Following *Sma*I-corII primer removal, 2,446,020 reads remained. After merging and trimming, 995,033 reads were retained and following dereplication, 755,865 reads remained across 12,145 ASVs. Following each analysis attempt, the total number of ASVs present in the output, the number of ASVs matching those in the custom database of putatively true Atlantic Whitefish ASVs, and the corresponding read depths of each were compared to determine analysis suitability (Table 10).

When analyzing the dilution series following the same downstream depth filtering protocols as the eDNA samples in section 3.3.3, only the five putatively true ASVs determined in section 3.4.2 were detected in the 10^{0} – 10^{-3} dilutions. In the 10^{-4} dilution, 27 ASVs were detected, including ASVs 01, 02, and 04, while the 10^{-5} dilution returned 10 ASVs, including ASV 01. The high variation in ASVs at the highest dilutions (lowest concentrations) in the series indicated that depth-based filtering to detect putatively true ASVs is unreliable for very low concentration eDNA samples.

The suitability of denoising and BLASTing against the custom reference database was explored as an alternative to depth-based ASV detection across the dilution series. BLASTing against the custom *Sma*I-corII reference database of putatively true ASVs (3.4.2) returned 341 unique sequences which had been assigned to one of the five *Sma*I-corII ASVs within the reference database. Of these returned sequences, 68 were assigned to ASV 01, 43 to ASV 02, 38 to ASV 03, 48 to ASV 04, and 34 to ASV 05. Most of the sequences were low depth and only 11 occurred in abundances greater than 1% depth at a given dilution. Seven of these sequences were only identified in the 10^{-3} – 10^{-5} dilutions, echoing the trend of increased variation at the highest dilutions (lowest concentrations) observed when using the depth-based filtering described above. In additional to the low depths of the returned sequences, the portion of the query sequence that matched, known as query coverage, varied widely among the 341 ASVs from 27% to 100%.

This indicated that while parts of the sequences aligned with the reference, the extent of alignment differed significantly from sequence to sequence.

Retaining ASVs with 100% query cover returned 33 sequences across the dilution series, however only six occurred in abundances greater than 1% at a given dilution: ASVs 01 – 05 and one additional sequence assigned to ASV 03 in the 10^{-4} dilution. Together, the low depths of most returned sequences suggest BLAST was unsuited for ASVs assessment even when using a custom reference database. The optimal analysis method was determined to be running the optimized custom pipeline and only retaining ASVs in the dereplicated output which were present in the custom *Sma*I-corII reference database of putatively true ASVs (3.4.2).

Table 10 Comparison of ASV results following application of alternative analysis methods to the tank dilution series.

Analysis description	**Number ASVs present in the custom *Sma*I-corII reference database**	**Total depth of ASVs in custom database**	**Total number of ASVs**	**Total depth of all ASVs**
Optimized custom pipeline (2.3.2) and downstream eDNA analysis (3.3.3)	5	152,015	38	167,756
Optimized custom pipeline (2.3.2) and only retained ASVs which were a 100% match to the custom reference database of *Sma*I-corII ASVs	5	157,117	5	157,117
Optimized custom pipeline, used the custom reference database of *Sma*I-corII ASVs when assigning taxonomy with BLAST (-perc_identity = 100)	5	157,117	341	178,283
As analysis 3, however only retained ASVs with 100% query coverage	5	157,117	33	159,135

In assessing the dilution series with the optimal method described in the previous paragraph, all five putatively true ASVs were detected across the dilution series in the 10^0 to 10^{-4} dilutions (Figure 11). In the 10^{-4} dilution, the pattern of ASV 01 comprising a higher percentage of detected reads disappeared; at that dilution, ASVs 01 and 04 were detected with similar read depth (1259 vs. 965, respectively). In the highest dilution of 10^{-5}, ASVs 01, 02, and 05 were detected with depths of 278, 27, and 1 reads, respectively. Across the dilution series, the combined read depth of all ASVs did not decrease in 10-fold increments: between the 10^0 and 10^{-1} dilution, there was a 21%, decrease in total read depth, a 13% decrease between 10^{-1} and 10^{-2} dilutions, a 2% decrease between 10^{-2} and 10^{-3} dilutions, a 91% decrease between 10^{-3} and 10^{-4} dilutions, and an 89% decrease between 10^{-4} and 10^{-5} dilutions.

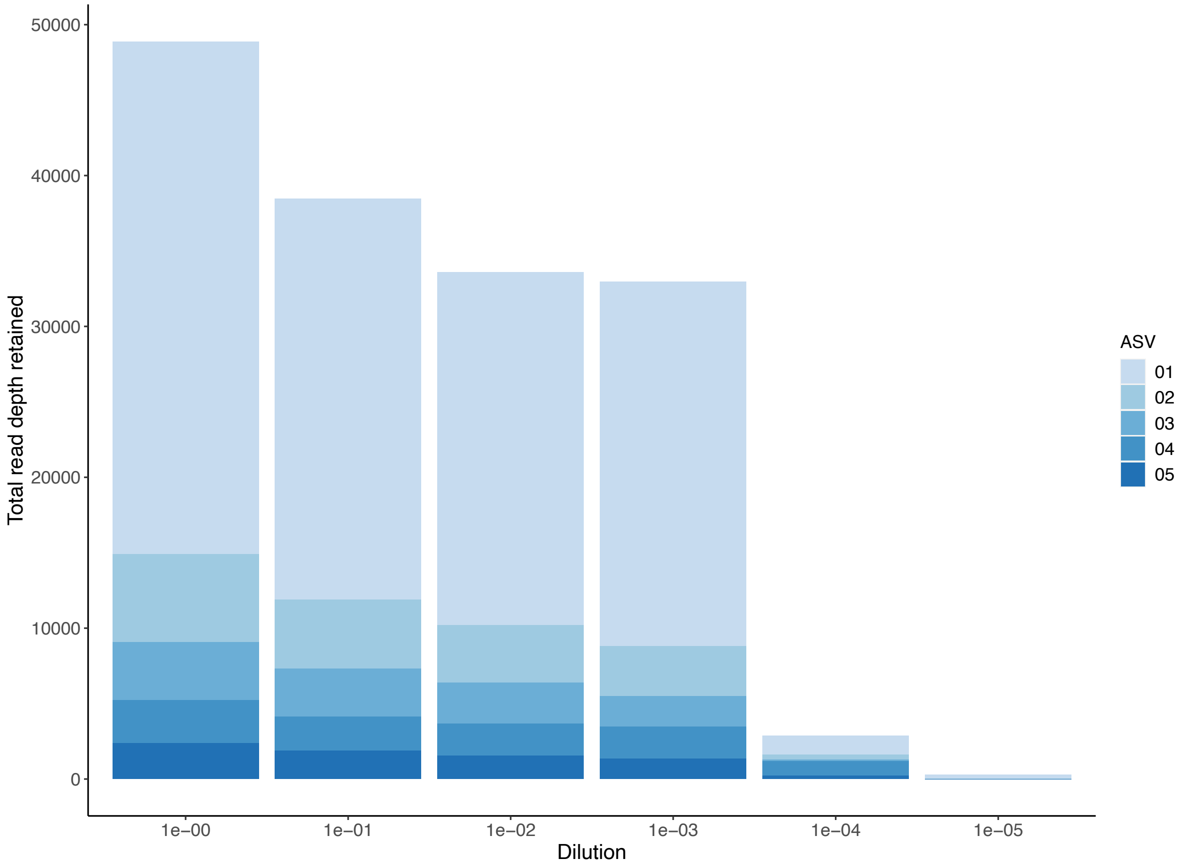

Figure 11 *Sma*I-corII ASV drop out across the tank eDNA dilution series.

3.5 Discussion

eDNA is most commonly used determine detection/non-detection of species within an environmental sample, and to a lesser extent, to estimate their abundance (Andres et al. 2023a), but there is growing interest in the use of eDNA tools for population-level analyses (Sigsgaard et al. 2016; Uchii et al. 2016; Goricki et al. 2017; Stat et al. 2017; Baker et al. 2018; Parsons et al. 2018; Marshall and Stepien 2019; Stepien et al. 2019; Jensen et al. 2020; Turon et al. 2020; Andres et al. 2021; Andres et al. 2023a,b). Here, I investigated sequence variation within the *Sma*I-corII SINE and examined the ability to detect these variants within Atlantic Whitefish eDNA samples. I demonstrated that commonly available bioinformatic pipelines must be validated and optimized for detection of variants within SINE amplicons obtained from both gDNA and eDNA samples. The *Sma*I-corII consensus sequence (ASV 01) comprised the majority of reads recovered in all Atlantic Whitefish samples assayed; however, four additional ASVs were consistently detected at lower but stable proportions. Analysis of *Sma*I-corII ASVs also identified variants unique to both Atlantic Whitefish and Lake Whitefish.

3.5.1 Bioinformatic pipeline comparison for *Sma*I-corII ASV analysis

Neither of two bioinformatic pipelines that are commonly used to analyze eDNA data, DADA2 and a custom UNIX-based pipeline, worked well for detecting variants of the *Sma*I-corII SINE in Atlantic Whitefish gDNA sequences when using the default parameters. The pipeline optimization undertaken in this study used retained read depth and number of ASVs consistently detected across PCR replicates of an individual fish as a proxy for pipeline suitability. Correction of erroneous sequences via DADA2's OMEGA_C parameter upwardly inflated retained read depth by 72%. Because the custom pipeline retained more consistently detected ASVs at similar depths to DADA2 with OMEGA_C parameter off, the custom pipeline was used for SINE ASV determination following optimization of its parameters. Taxonomy assignment via BLAST resulted in a large proportion of reads being dropped. Depth-based filtering on the dereplicated dataset was used to generate a custom database of Atlantic Whitefish ASVs which was then used to examine the number of detectable ASVs in field samples. These results show the importance of validating bioinformatic tools when applying them beyond their original applications, as the targeted nature of this SINE did not justify this method of assignment.

3.5.2 Intraspecific variation of *Sma*I-corII

Understanding of intraspecific variation is essential for effective monitoring and management of aquatic populations (Andres et al. 2023a). Assessing population-level genetic variation using eDNA is attractive because it enables detection and monitoring of population dynamics without need for direct observation or handling. Five putatively true ASVs were identified from Atlantic Whitefish gDNA samples, four of which were found in all individuals (Figure 8). The identification of ASV 05 in all but two of the individuals indicates variability among individual Atlantic Whitefish in their *Sma*I-corII repeat copies. As the ASV identification undertaken in this study was heavily influenced by initial sequencing depth, gDNA samples from additional Atlantic Whitefish should be sequenced at higher depths to further confirm the level of intraspecific variation present.

Atlantic Whitefish occur in three lakes in the Petite Rivière watershed (Bradford et al. 2004). Individuals within this system are considered to belong to a single population of unknown but likely small size (Murray 2005; Cook 2012). Analysis of 15 microsatellite markers in 116

Atlantic Whitefish from the Petite Rivière system revealed remarkably low genetic diversity compared to populations of Lake Whitefish and Cisco (*Coregonus artedi*), two closely related species: of the 12 loci which amplified across all three species (one of which did not amplify in Cisco), 33% were polymorphic in Atlantic Whitefish compared to 83% and 82% in Lake Whitefish and Cisco, respectively (Murray 2005). Atlantic Whitefish also showed lower unbiased allelic richness (A_E) and expected heterozygosity (H_E) at 1.39 and 0.14, respectively, compared to Lake Whitefish (A_E = 2.51, H_E =0.38) and Cisco (A_E = 3.83, H_E =0.56) (Murray 2005). Further analysis of the 15 microsatellite loci in 169 Atlantic Whitefish confirmed the low genetic diversity observed within the species with an average observed heterozygosity of 0.27 (Cook 2012). Despite the low genetic diversity of the Petite Rivière population of Atlantic Whitefish, a test for a recent bottleneck produced a non-significant result (Murray 2005; Cook 2012). Later analysis of 97 Atlantic Whitefish mitochondrial genomes revealed only 13 haplotypes, 12 of which differ from the most common haplotype by 1–2 SNPs, and most of which were observed in 1–3 individuals (Einfeldt et al. *in prep*). Therefore, identification of variant mitochondrial haplotypes from eDNA samples would be challenged both by the rareness of individuals with the variants and the need for multiple targeted, haplotype-specific assays.

eDNA analyses face limitations due to the positive relationship between species abundance and eDNA concentration, particularly challenging when targeting rare species or variant alleles/haplotypes occurring at low abundances (Andres et al. 2023b). Targeting high copy sequences, such as TEs, can help offset the challenges of estimating genetic diversity within a target species via eDNA analysis due to its increased abundance in a system compared to mitochondrial markers. Understanding the level of genetic diversity is crucial for assessing a species' current health and its potential for survival (Hughes et al. 2008). Loss of genetic variability can be detrimental to a population as increased homozygosity may diminish its ability to adapt or respond to habitat changes. In the case of single population species, such as Atlantic Whitefish, a decline in the population could potentially lead to extinction. Therefore, it is crucial to comprehend and monitor the level of genetic diversity present. The detection of a TE variant in the single, genetically depauperate Atlantic Whitefish population suggests segregating variants may be present in other, more abundant species with multiple populations. Future work should examine whether population mixing can be detected through eDNA analysis by targeting TEs.

3.5.3 Interspecific variation of *Sma*I-corII

Analysis of eDNA from a lake containing Lake Whitefish revealed both similarities and differences in *Sma*I-corII variants between Atlantic Whitefish and Lake Whitefish. Of the shared ASVs between the species, the relative abundance of the ASVs differed between Lake Whitefish and the Atlantic Whitefish samples. Although ASV 01 (*Sma*I-corII consensus) was present as the most abundant sequence in both species, it accounted for a significantly larger proportion of reads in the Atlantic Whitefish samples (avg. 69%) than the Lake Whitefish samples (59%); However, direct comparisons between the complete genome for the two species should be made to compare the abundance of each ASV.

The presence of four shared ASVs between Atlantic Whitefish and Lake Whitefish suggests that these variants may predate the divergence of the species while the two ASVs only present in the Lake Whitefish samples (ASVs 06 and 07) and the ASV only present in the Atlantic Whitefish samples (ASV 05) evolved later, however further samples from each species should be evaluated. Three of the seven identified ASVs were differentiated by the insertion of one or more adenine bases and this run of adenine bases was observed by Hamada et al. (1997) in the other coregonid species examined during the initial *Sma*I-corII discovery (Figure 10). These results suggest the specificity of the SINE may be increased to the species-level through ASV identification and the signal of these species-specific ASVs may be comparable to what would be expected from a mtDNA marker given the high copy number of the SINE marker; however further investigation of Lake Whitefish gDNA samples is required to confirm the unique ASVs detected.

3.5.4 eDNA, TEs, and population genetics

Incorporation of reliable population-level inferences into eDNA analysis would introduce an additional non-invasive tool for the management of aquatic populations (Parsons et al. 2018). Multiple mitochondrial D-loop haplotypes have been detected in eDNA samples for several species. For Whale Shark (Sigsgaard et al. 2016) and harbour porpoise (*Phocoena phocoena*) (Parsons et al. 2018), the relative abundance of haplotypes detected from eDNA analysis corresponded to the population frequencies known from conventional genetic sampling of multiple individuals. eDNA based detection of D-loop haplotypes was also used to differentiate between Killer Whale (*Orcinus orca*) ecotypes (Baker et al. 2018). Detection of multiple haplotypes within a single eDNA sample would allow for estimates of haplotype diversity, nucleotide diversity, and segregating sites for a population (Andres et al. 2023a). Although detection of genetic variation at single-copy nuclear markers is more challenging than targeting mtDNA, through sequencing a panel of multiallelic microsatellite markers, Andres et al. (2023b) successfully detected patterns of allele frequencies and genetic variability within and among populations of invasive Round Goby across their invaded range within the Great Lakes. Of the microsatellites targeted, concentrations of nuDNA markers were lower than mtDNA (Andres et al. 2023b). Estimates of nuDNA allele frequencies from eDNA samples could enable estimates of population structure via assessment of how allelic richness, expected heterozygosity, and the number of private alleles vary among populations (Andres et al. 2023a) as well as fixation index or other genetic distance measures based on allele frequencies. Another possibility is relating genetic diversity estimates to species abundance to bypass the challenges of estimating abundance from eDNA concentration to abundance (Andres et al. 2023a). When concentrations of Round Goby mitochondrial eDNA (mt-eDNA) and nu-eDNA were compared in a field setting, mt-eDNA was detected at significantly higher abundances with an average ratio of mt-eDNA:nu-eDNA of 196:1, though there was large variation in this ration between sites (20:1 to 1132:1) (Andres et al. 2023b).

This research represents the first application of TEs in eDNA analysis. However, TEs, including SINEs, are ubiquitous in eukaryotic species, and the results presented herein demonstrate the feasibility of detecting intraspecific SINE variation in a low-abundance species. When choosing TEs for assay development, their suitability should be validated on a per-species basis as particular families of TEs do not occur in a consistent copy number across species. For example, *Hpa* elements, though present in all salmonids, are 5-fold and 20-fold to 200-fold less abundant in graylings (Thymallinae) and whitefish (Coregoninae) species compared to Salmoninae species (huchens, trout, char, salmon), respectively (Kido et al. 1994). The *Ava*III SINE is also present across salmonid species; however, its abundance within the host genome is low, at approximately 10^2 copies/cell (Kido et al. 1994), which may negate the advantages of SINE-based eDNA approaches over mtDNA.

Although not previously targeted in eDNA studies, TEs have previously been examined in population genetic studies using gDNA, though to a lesser degree than SNPs (Bourgeois and Boissinot 2019). TE polymorphisms can take the form of overall copy number variation (Stapley et al. 2015), sequence variation within a particular type of TE (Kramerov and Vassetzky 2011), and variation in where copies of TEs are inserted in the genome (Bourque et al. 2018), and many studies have focused on the latter in population genetic studies. The first models for TE polymorphism were developed in the 1980s to assess how distributions of TEs differed within diploid, sexually reproducing populations of *Drosophila melanogaster* mating at random (Charlesworth and Charlesworth 1983). Owing to the dispersed and irreversible nature of SINE insertions across the genome, polymorphic human *Alu* elements were found to be well suited for population genetic and phylogenetic studies (Kothe et al. 2016); assessing the presence/absence of *Alu* polymorphic elements at different insertion sites the across 26 human populations were found to reflect known patterns of human evolution (Rishishwar et al. 2015). Chen et al. (2021) successfully analyzed 30 SINE retrotransposon insertion polymorphisms markers across seven Chinese miniature pig populations comprising multiple pig species to determine genetic diversity, differentiation, and structure between the population.

Current eDNA analyses commonly focus on short amplicons (100 – 400 bp) (Adams et al. 2019) which could limit the usefulness of TE's using an eDNA approach for intraspecific analysis. This limitation arises as both the TE and the adjacent genes would need to be amplified to assess the insertion locations. In a 2019 review discussing the current state of eDNA population genetic analyses, Adams et al. (2019) proposed the potential benefits of sequencing longer eDNA fragments to achieve more comprehensive mtDNA haplotype resolution. Using long-range PCR, Deiner et al. (2017) were able to sequence mitochondrial genomes (>16 kb) from eDNA samples and it is conceivable that targeting longer nuDNA fragments encompassing TEs and the surrounding regions might enable intraspecific TE analysis, however further testing is necessary to ascertain the feasibility of implementing such a method as longer fragments degrade quicker within the environment (Bylemans et al. 2018). To date, assessing TE sequence variations, as demonstrated in this study, may be the most feasible method of using TEs to determine intraspecific diversity in an eDNA study.

3.5.5 Challenges of ASV detection with decreasing eDNA concentrations

In eDNA analysis, the concentration of eDNA within the sample cannot be controlled and therefore it is vital to understand how ASV detections are influenced by eDNA concentration. The dilution series highlights that while Atlantic Whitefish *Sma*I-corII variation was detected in eDNA samples, detection of variants decreased with decreasing sample concentration. Across the dilution series, ASV 01 occupied the highest proportion of reads per dilution, consistent with other ASV analyses undertaken in this study. Interestingly, the ASVs did not drop out of the dilution series in proportion to their sequence depths in the 10^0 sample; the fifth most abundant ASV, 05 (4.83% of 10^0 depth), was detected as a singleton in the highest dilution of 10^{-5} whereas ASVs 03 (7.89%) and 04 (5.88%) dropped out in the 10^{-4} dilution. The detection of 2/5 ASVs (01 and 02) in the highest dilution (10^{-5}) at depths greater than 1 read highlights the sensitivity of this marker as the limit of *Sma*I-corII variant drop out was not determined in this series. Across the dilution series, the retained read depth experienced a sharp decline between the 10^{-3} and 10^{-4} dilutions and a finer resolution dilution series concentrated on this range should be explored (Figure 11).

3.5.6 Limitations and biases

Along with the indexing error which led to *Lobelia* spp. sequences skewing the count of input reads (Appendix F), this study has additional limitations. Notably, this study only compared two pipelines out of the many that are available (Mathon et al. 2021) for *Sma*I-corII ASV assessment and used a small sample size of 16 fish and eleven 1 L eDNA samples from three different sources. Further, the gDNA sequences used for determination of putatively true *Sma*I-corII ASVs underwent single-end sequencing, and the merging step of the pipeline was omitted. Although analysis of the paired-end-sequenced eDNA samples included merging of the forward and reverse reads and resulted in the detection of all five ASVs identified in the gDNA samples, confidence in the gDNA results will be further strengthened by paired-end sequencing and merging optimization. Similarly, Lake Whitefish gDNA should be sequenced to confirm the species-specific variants (ASVs 06 and 07) identified in the Shingle Lake eDNA samples.

Assumptions were made during the custom pipeline optimization which could have biased the results. While not every individual was expected to contain every *Sma*I-corII variant, true ASVs were expected to be consistently detected across the technical PCR replicates of individual fish (gDNA samples) and this assumption was used as to filter through the thousands of ASVs returned in the post-dereplicated output. As depth-based thresholds were used, ASV identification was influenced by the sequencing depth obtained during Illumina sequencing and Smith and Peay (2014) recommend increasing sequencing depth rather than technical replicates to improve ecological inferences from NGS data.

The ASV determination undertaken in this study involved a trade-off between removing erroneous sequences and detecting true variants (Couton et al. 2021) which was further complicated by the unknown mutation rate of the *Sma*I-corII SINE within the Atlantic Whitefish genome. The presence of ASVs 01 – 04 in both Atlantic Whitefish and Lake Whitefish suggest the formation of these variants occurred prior to the divergence of the two species and have persisted within the respective genomes for over 28 MY divergence-time (Crête-Lafrenière et al. 2012). Applying 1% depth cut-offs across all PCR replicates and on a per-ASV basis may have resulted in the elimination of some true but rare *Sma*I-corII variants in the final output, however stringent filtering criteria was employed to reduce the over-estimation of false-positive variants (Lerch et al. 2017). The removal of rare ASVs or operational taxonomic units (OTUs; clusters of ASVs which are 97% similar to one another) during bioinformatic processing of NGS data is not uncommon (Brown et al. 2015), however, there is no standardized threshold (Pauvert et al. 2019); Zanovello et al. (2023) applied a 5% OTU depth cut-off, Andres et al. (2021) applied a 1% allele frequency cut-off (similarly to the 1% ASV cut-off applied in this study), and Andres et al. (2023b) removed alleles with fewer than 2 reads in a sample to account for low levels of contamination. Ultimately, as eDNA is highly susceptible to contamination and there is growing interest in incorporating eDNA-derived results in management plans, using conservative thresholds during analysis helps to maintain confidence that results are accurate while acknowledging the possibility of false-negatives. Future simulation analyses of TE mutation rate may help to better define appropriate thresholds for detection of real but rare TE variants.

3.5.7 Conclusions

The non-invasive nature of eDNA sampling coupled with its relative ease of use and reduced cost compared to traditional aquatic monitoring methods has generated much interest in applying eDNA analysis to management conservation efforts, particularly for species which are at-risk and/or elusive (Adams et al. 2019). The detection of intraspecific genetic variation from eDNA samples will expand the limits of this tool beyond its most common applications of species detection/non-detection for effective management of aquatic populations (Andres et al. 2023a) by providing insights regarding gene flow, genetic drift, mutation, and natural selection (Adams et al. 2019). Additionally, it may be possible to identify an invasive species source population and invasion pathways (Jerde et al. 2011; Adams et al. 2019). Through optimization of commonly used bioinformatic tools, this work highlights the ability to determine SINE variants from gDNA and apply this reference database to eDNA samples, providing an additional resource within the eDNA toolbox.

Chapter 4 – Conclusion

4.1 Summary

As an emerging tool, eDNA methodology lacks standardization across studies, and numerous questions persist regarding the reliability of information obtainable through eDNA analyses (Loeza-Quintana et al. 2020). Despite variations in the application of eDNA tools, the majority of studies primarily focus on mtDNA due, in part, to its advantageous copy number per cell abundance compared to most nuclear markers (Jo et al. 2019). Here, I developed, validated, and compared a novel eDNA marker targeting a highly repetitive SINE with a conventional mtDNA marker, using Atlantic Whitefish as a test case. Atlantic Whitefish, an endangered member of the salmonid Coregoninae subfamily, inhabit only three connected lakes covering 16 km^2 of surface area (Bradford et al. 2004). Within this system, they remain elusive and challenging to detect using traditional aquatic monitoring methods. Currently, Atlantic Whitefish are being raised in captivity at Dalhousie's Aquatron Facility as part of a species recovery strategy. Their endangered status, limited distribution, and captive population render Atlantic Whitefish an ideal candidate for developing an alternative eDNA marker, aiming to enhance the ability to detect low-abundance species through eDNA analyses.

In Chapter 2, I developed and validated qPCR assays targeting the *Sma*I-corII SINE and the mitochondrial ND4 subunit. To determine the limits of each assay, the LOD and LOQ were determined by assessing multiple replicates of a dilution series of synthetic DNA fragments. Both assays had equal LOQs of 64 c/µL while the LOD of the ND4 assay was four times lower than that of *Sma*I-corII, indicating this assay has an increased ability to detect target molecules within a sample. Across all direct comparisons of ND4 and *Sma*I-corII during assay validation, *Sma*I-corII amplified an average of 7.2 cycles before ND4 which corresponds *Sma*I-corII copies occurring in abundances an average of 146x greater than ND4. When the assays were then applied to field samples collected from a net pen housing 150 juvenile Atlantic Whitefish in Milipsigate Lake, roughly 100x more copies of *Sma*I-corII were present within the net pen than ND4 (Figure 2). Despite its lower LOD, the 20 m distance of detection of the ND4 assay was four times smaller than that of the *Sma*I-corII assay. eDNA yield from different densities of juvenile Atlantic Whitefish were also assessed to determine if yield scales proportionally to density. Though eDNA yield for *Sma*I-corII did not scale as expected (Figure 3), copies of *Sma*I-corII were again detected in abundances magnitudes of order greater than ND4. The results of the field samples support the findings that compared to ND4, the high copy number of *Sma*I-corII targets within a given sample lead to an increased detection ability of Atlantic Whitefish eDNA.

In Chapter 3, I showcased that SINEs, in addition to enhancing detection capabilities, can provide intraspecific insights through eDNA analysis. To evaluate the most suitable pipeline for analyzing *Sma*I-corII ASVs considering the inherent variability of SINEs and non-identical copies throughout the genome (Kramerov and Vassetzky 2011), I compared the DADA2 pipeline with a custom UNIX-based pipeline. DADA2's core algorithm was found to be less appropriate for *Sma*I-corII ASV determination compared to the custom pipeline. This was primarily due to the error estimation and correction functions in DADA2, which retained more erroneous sequences. Disabling the error correction parameter (OMEGA_C) substantially reduced the number of retained reads, and optimization efforts were concentrated on the custom pipeline, which was initially designed for metabarcoding purposes.

The optimal approach for analyzing *Sma*I-corII sequences involved running the custom pipeline up to the dereplication stage and then applying depth-based cut-offs for each PCR replicate and ASV. These thresholds enabled the detection of five ASVs in Atlantic Whitefish gDNA samples, four of which were identified in all examined individuals. This finding suggests that, despite their inferred extremely low population size (Cook 2012), Atlantic Whitefish maintain intraspecific genetic diversity within the *Sma*I-corII repeats. The five ASVs were confirmed in eDNA samples collected from both the Aquatron and the net pen in Milipsigate Lake.

Additionally, I demonstrated that ASV analysis can be used to identify species-specific *Sma*I-corII ASVs by identifying one ASV unique to Atlantic Whitefish samples and two ASVs unique to Lake Whitefish samples.

4.2 <u>Applications for conservation</u>

Biodiversity is in crisis, with species loss accelerating due to anthropogenic factors (Barnosky et al. 2011). The effects of biodiversity loss are cascading and far-reaching. While it has been observed across all habitat types, freshwater vertebrates have experienced a decline twice the rate of those in marine or terrestrial ecosystems (Almond et al. 2022). Management of species at-risk requires accurate and reliable methods for detecting and monitoring species within their habitats and the non-invasive nature of eDNA tools have led to interest in its inclusion along other aquatic conservation efforts (Bernos et al. 2023).

Here I demonstrated that targeting a highly abundant TE led to increased detection of a rare species via eDNA analysis compared to a mitochondrial marker. As species and their associated eDNA concentrations decline, the number of target molecules in the environment will decline in tandem and targeting a highly abundant marker such as TEs may at least partially offset this technical challenge. The increased sensitivity of TEs will be particularly advantageous for eDNA analyses of endangered species which typically occur in low abundances and can be difficult to detect with conventional eDNA markers, as demonstrated in this study. In previous eDNA studies, successful detection of spawning events using both mtDNA and nuDNA markers has been reported (Bylemans et al. 2017; Tillotson et al. 2018). Nonetheless, TEs could also offer advantages for this application, especially when dealing with species occurring in extremely low abundances or when the reproductive event produces a small signal that conventional markers might miss. As the SINE marker used in this study was able to detect Atlantic Whitefish eDNA up to 80 m from source, this study also demonstrated that efforts should be taken to determine the distance of detection of TE markers to ensure eDNA detection results can be reliably related to species presence within a given area. The increased distance of detection of the SINE marker will also be valuable for early detection of invasive species, allowing mitigation efforts to begin before invasion pathways are fully established as once a species is established, complete eradication efforts may not be possible (Lodge et al. 2006). Within eDNA studies, the applications of TEs will likely be limited to targeted detections. This limitation arises from the variable nature of TEs, which poses challenges in designing a universal primer for amplifying the DNA of most species via metabarcoding. Therefore, TE assays may be used to supplement existing metabarcoding assays where the eDNA of more abundant species can mask the eDNA of less abundant species within a sample (Evans et al. 2016).

Beyond their increased detection abilities, TEs are a source of genetic diversity and intraspecific variation was observed within Atlantic Whitefish despite their inferred small population size (Cook 2012). The tank dilution series of this study highlighted that even at very low concentrations, TE variants can be detected from eDNA samples. TEs are present within nearly all eukaryotic genomes (Kramerov and Vassetzky 2011) and are well-studied in certain commercially and culturally important species, such as salmonids (Matveev and Okada 2009). Future research should explore the potential of TEs in species with multiple populations, both in marine and freshwater environments, to determine whether eDNA analysis of TEs can differentiate between these populations, in addition to their enhanced detection capabilities.

Meeting the recovery objectives for Atlantic Whitefish requires the use of sensitive and accurate monitoring techniques within their natural habitat. The *Sma*I-corII marker exhibited heightened sensitivity in the detection of Atlantic Whitefish compared to a conventional mitochondrial maker. Nevertheless, it is essential to conduct further analysis to ascertain the marker's limitations, as consistent CPs were observed across net pen-associated samples (Figure 2), and the read depth in the tank dilution series remained steady between 10^0 and 10^{-3} dilutions before a sharp decline (Figure 11). The improved sensitivity offered by TE markers can serve not only in the detection of Atlantic Whitefish but also in identifying their aquatic invasive species predators, such as Smallmouth Bass and Chain Pickerel. If Atlantic Whitefish are to be introduced to new locations as outlined in their recovery plan by DFO (2018), the heightened sensitivity of TE markers can ensure the suitability of the habitat prior to their release. In the context of the Petite Rivière watershed, the eDNA tools developed in this study can be broadly applied to locate spawning activity, addressing a critical gap in our understanding.

4.3 **Final thoughts**

In many ways, eDNA analysis is still an emerging tool and there remain knowledge gaps surrounding associated technical challenges. Assay development of the markers used within this study was a non-linear process that occurred over many months and reanalysis of samples was not possible due to degradation of the extracts. Further work should examine best storage practices for eDNA samples to allow for sample archiving and reanalysis as needed, ensuring the data remains accessible over time. Limitations aside, this study demonstrated that high copy nuclear repeats offer increased detection of a low abundance freshwater fish species and these markers may be applied to other species. Overall, these results highlight the importance of exploring alternative markers in eDNA analysis to continually increase the resolution of eDNA results, thereby enhancing its utility in aquatic conservation.

References

Abbott C, Coulson M, Gagné N, Lacoursière-Roussel A, Parent GJ, Bajno R, Dietrich C, May-McNally S. 2021. Guidance on the Use of Targeted Environmental DNA (eDNA) Analysis for the Management of Aquatic Invasive Species and Species at Risk. DFO Can. Sci. Advis. Sec. Res. Doc. 2021/019. iv + 42 p.

Adams CIM, Knapp M, Gemmell NJ, Jeunen GJ, Bunce M, Lamare MD, Taylor HR. 2019. Beyond biodiversity: Can environmental DNA (eDNA) cut it as a population genetics tool? Genes (Basel). 10(3). doi:10.3390/genes10030192.

Akbari M, Hansen MD, Halgunset J, Skorpen F, Krokan HE. 2005. Low Copy Number DNA Template Can Render Polymerase Chain Reaction Error Prone in a Sequence-Dependent Manner. J Mol Diagn. 7(1):36. doi:10.1016/S1525-1578(10)60006-2.

Alexander JB, Marnane MJ, McDonald JI, Lukehurst SS, Elsdon TS, Simpson T, Hinz S, Bunce M, Harvey ES. 2023. Comparing environmental DNA collection methods for sampling community composition on marine infrastructure. Estuar Coast Shelf Sci. 283:108283. doi:10.1016/J.ECSS.2023.108283.

Almond REA, Grooten M, Juffe Bignoli D, Petersen T, editors. 2022. Living Planet Report 2022 – Building a nature-positive society. WWF, Gland, Switzerland.

Andres KJ, Lodge DM, Andrés J. 2023b. Environmental DNA reveals the genetic diversity and population structure of an invasive species in the Laurentian Great Lakes. Proc Natl Acad Sci U S A. 120(37):e2307345120.

Andres KJ, Lodge DM, Sethi SA, Andrés J. 2023a. Detecting and analysing intraspecific genetic variation with eDNA: From population genetics to species abundance. Mol Ecol. 32(15):4118–4132. doi:10.1111/MEC.17031.

Andres KJ, Sethi SA, Lodge DM, Andrés J. 2021. Nuclear eDNA estimates population allele frequencies and abundance in experimental mesocosms and field samples. Mol Ecol. 30(3):685. doi:10.1111/MEC.15765.

Andrews S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Baillie SM, McGowan C, May-McNally S, Leggatt R, Sutherland BJG, Robinson S. 2019. Environmental DNA and its applications to Fisheries and Oceans Canada: National needs and priorities. Can. Tech. Rep. Fish. Aquat. Sci. 3329: xiv + 84 p.

Baker CS, Steel D, Nieukirk S, Klinck H. 2018. Environmental DNA (eDNA) from the wake of the whales: Droplet digital PCR for detection and species identification. Front Mar Sci. 5:1–11. doi:10.3389/fmars.2018.00133.

Barnes MA, Turner CR. 2016. The ecology of environmental DNA and implications for conservation genetics. Conserv Genet. 17(1):1–17. doi:10.1007/s10592-015-0775-4.

Barnosky AD, Matzke N, Tomiya S, Wogan GOU, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, et al. 2011. Has the Earth's sixth mass extinction already arrived? Nature. 471(7336):51–57. doi:10.1038/nature09678.

Beng KC, Corlett RT. 2020. Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. Biodivers Conserv. 29(7):2089–2121. doi:10.1007/S10531-020-01980-0.

Bernos TA, Yates MC, Docker MF, Fitzgerald A, Hanner R, Heath D, Imrit A, Livernois J, Myler E, Patel K, et al. 2023. Environmental DNA (eDNA) applications in freshwater fisheries management and conservation in Canada: overview of current challenges and opportunities. Can. J. Fish. Aquat. Sci. 80(7):1170–1186. doi.org/10.1139/cjfas-2022-0162.

Biémont C, Vieira C. 2006. Junk DNA as an evolutionary force. Nature 2006 443:7111. 443(7111):521–524. doi:10.1038/443521a.

Blackman RC, Walser JC, Rüber L, Brantschen J, Villalba S, Brodersen J, Seehausen O, Altermatt F. 2023. General principles for assignments of communities from eDNA: Open versus closed taxonomic databases. Environ DNA. 5(2):326–342. doi:10.1002/EDN3.382.

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods 2012 10:1. 10(1):57–59. doi:10.1038/nmeth.2276.

Bourgeois Y, Boissinot S. 2019. On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. Genes. 10(6):419. doi:10.3390/GENES10060419.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al. 2018. Ten things you should know about transposable elements. Genome Biol. 19(1):1–12. doi:10.1186/S13059-018-1577-Z.

Bradford R, Longard DA, Longue P. 2004. Status, trends, and recovery considerations in support of an allowable harm assessment for Atlantic Whitefish (*Coregonus huntsmani*). DFO Can. Sci. Advis. Sec. Res. Doc. 2004/109.

Bradford RG, CMB Mahaney. 2004. Distributions of lake whitefish (*Coregonus clupeaformis*) fry from the lower Great Lakes federal hatcheries to elsewhere in Canada and beyond (Years 1878-1 91 4). Can. Data Rep. Fish. Aquat. Sci. 1 149: iii + 19 p.

Brown SP, Veach AM, Rigdon-Huss AR, Grond K, Lickteig SK, Lothamer K, Oliver AK, Jumpponen A. 2015. Scraping the bottom of the barrel: are rare high throughput sequences artifacts? Fungal Ecol. 13:221–225. doi:10.1016/J.FUNECO.2014.08.006.

Brys R, Haegeman A, Halfmaerten D, Neyrinck S, Staelens A, Auwerx J, Ruttink T. 2021. Monitoring of spatiotemporal occupancy patterns of fish and amphibian species in a lentic aquatic system using environmental DNA. Mol Ecol. 30(13):3097–3110. doi:10.1111/MEC.15742.

Burns M, Valdivia H. 2008. Modelling the limit of detection in real-time quantitative PCR. Eur Food Res and Technol. 226(6):1513–1524. doi:10.1007/S00217-007-0683-Z/FIGURES/8.

Bylemans J, Furlan EM, Gleeson DM, Hardy CM, Duncan RP. 2018. Does Size Matter? An Experimental Evaluation of the Relative Abundance and Decay Rates of Aquatic Environmental DNA. Environ Sci Technol. 52(11):6408–6416. doi:10.1021/acs.est.8b01071.

Bylemans J, Furlan EM, Hardy CM, McGuffie P, Lintermans M, Gleeson DM. 2017. An environmental DNA-based method for monitoring spawning activity: a case study, using the endangered Macquarie perch (*Macquaria australasica*). Methods Ecol Evol. 8(5):646–655. doi:10.1111/2041-210X.12709.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 13(7):581–583. doi:10.1038/nmeth.3869.

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, MacE GM, Tilman D, Wardle DA, et al. 2012. Biodiversity loss and its impact on humanity. Nature. 486(7401):59–67. doi:10.1038/NATURE11148.

Casacuberta E, González J. 2013. The impact of transposable elements in environmental adaptation. Mol Ecol. 22(6):1503–1517. doi:10.1111/MEC.12170.

Cha RS, Thilly WG. Specificity, Efficiency, and Fidelity of PCR. Genome Res. 3: S18–S29.

Charlesworth B, Charlesworth D. 1983. The population dynamics of transposable elements. Genet Res, Camb. 42:1–27. doi:10.1017/S0016672300021455.

Chen C, Wang X, Zong W, D'alessandro E, Giosa D, Guo Y, Mao J, Song C. 2021. Genetic diversity and population structures in Chinese miniature pigs revealed by SINE retrotransposon insertion polymorphisms, a new type of genetic markers. Animals. 11(4). doi:10.3390/ani11041136.

Cook AM. 2012. Addressing key conservation priorities in a data poor species. Ph.D. thesis. Dalhousie University, Halifax, Nova Scotia. 198pp.

COSEWIC. 2010. COSEWIC assessment and status report on the Atlantic Whitefish *Coregonus huntsmani* in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. x + 31 pp.

Couton M, Baud A, Daguin-Thiébaut C, Corre E, Comtet T, Viard F. 2021. High-throughput sequencing on preservative ethanol is effective at jointly examining infraspecific and taxonomic diversity, although bioinformatics pipelines do not perform equally. Ecol Evol. 11(10):5533–5546. doi:10.1002/ECE3.7453.

Crête-Lafrenière A, Weir LK, Bernatchez L. 2012. Framing the Salmonidae Family Phylogenetic Portrait: A More Complete Picture from Increased Taxon Sampling. PLoS One. 7(10). doi:10.1371/journal.pone.0046662.

Deiner K, Renshaw MA, Li Y, Olds BP, Lodge DM, Pfrender ME. 2017. Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. Methods Ecol Evol. 8(12):1888–1898. doi:10.1111/2041-210X.12836.

Department of Fisheries and Oceans. 2006. Recovery Strategy for the Atlantic Whitefish (*Coregonus huntsmani*) in Canada. Species at Risk Act Recovery Strategy Series. Fisheries and Oceans Canada, Ottawa, xiii + 42 pp.

Department of Fisheries and Oceans Canada. 2018. Recovery strategy for the Atlantic Whitefish (*Coregonus huntsmani*) in Canada. Species at Risk Act recovery strategy Series. Fisheries and Oceans Canada, Ottawa. xiii + 62 pp.

Dudgeon D. 2010. Prospects for sustaining freshwater biodiversity in the 21st century: linking ecosystem structure and function. Curr Opin Environ Sustain. 2(5–6):422–430. doi:10.1016/J.COSUST.2010.09.001.

Dudgeon D, Arthington AH, Gessner MO, Kawabata ZI, Knowler DJ, Lévêque C, Naiman RJ, Prieur-Richard AH, Soto D, Stiassny MLJ, et al. 2006. Freshwater biodiversity: Importance, threats, status and conservation challenges. Biol Rev Camb Philos Soc. 81(2):163–182. doi:10.1017/S1464793105006950.

Dunker KJ, Sepulveda AJ, Massengill RL, Olsen JB, Russ OL, Wenburg JK, Antonovich A. 2016. Potential of Environmental DNA to Evaluate Northern Pike (*Esox lucius*) Eradication Efforts: An Experimental Test and Case Study. PLoS One. 11(9):e0162277. doi:10.1371/JOURNAL.PONE.0162277.

Dysthe JC, Franklin TW, McKelvey KS, Young MK, Schwartz MK. 2018. An improved environmental DNA assay for bull trout (*Salvelinus confluentus*) based on the ribosomal internal transcribed spacer I. PLoS One. 13(11):e0206851. doi:10.1371/JOURNAL.PONE.0206851.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792-97.

Edgar RC. 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv.:081257. doi:10.1101/081257.

Edgar RC. 2016. UCHIME2: improved chimera prediction for amplicon sequencing. bioRxiv.:074252. doi:10.1101/074252.

Edge TA. 1984. Preliminary Status of the Acadian Whitefish, *Coregonus Canadensis*, in Southern Nova Scotia. Canadian field-naturalist 98(1):86–90.

Edge TA, Gilhen J. 2001. Draft. Update COSEWIC status report on Atlantic whitefish, *Coregonus huntsmani*. Prepared for the Committee on the Status of Endangered Wildlife in Canada (COSEWIC), Canadian Wildlife Service, Ottawa, Ontario. September 12, 2001. 47p. + Tables.

Einfeldt T, Watson B, Paterson I, Broome J, Bentzen P. *in prep* Extreme low genetic diversity and interannual variation in basal group of coregonid radiation, the critically endangered Atlantic Whitefish (*Coregonus huntsmani*).

Elbarbary RA, Lucas BA, Maquat LE. 2016. Retrotransposons as regulators of gene expression. Science. 351(6274):aac7247. doi: 10.1126/science.aac7247.

Enns A, Kraus D, Hebb A. 2020. Ours to Save. NatureServe Canada and Nature Conservancy of Canada.

Evans NT, Olds BP, Renshaw MA, Turner CR, Li Y, Jerde CL, Mahon AR, Pfrender ME, Lamberti GA, Lodge DM. 2016. Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Mol Ecol Resour. 16(1):29–41. doi:10.1111/1755-0998.12433.

Fediajevaite J, Priestley V, Arnold R, Savolainen V. 2021. Meta-analysis shows that environmental DNA outperforms traditional surveys, but warrants better reporting standards. Ecol Evol. 11(9):4803–4815. doi:10.1002/ECE3.7382.

Ficetola GF, Miaud C, Pompanon F, Taberlet P. 2008. Species detection using environmental DNA from water samples. Biol Lett. 4(4):423–425. doi:10.1098/RSBL.2008.0118.

Foran DR. 2006. Relative Degradation of Nuclear and Mitochondrial DNA: An Experimental Approach. J Forensic Sci. 51(4):766–770. doi:10.1111/J.1556-4029.2006.00176.X.

Forootan A, Sjöback R, Björkman J, Sjögreen B, Linz L, Kubista M. 2017. Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR). Biomol Detect Quantif. 12:1–6. doi:10.1016/J.BDQ.2017.04.001.

Gao B, Shen D, Xue S, Chen C, Cui H, Song C. 2016. The contribution of transposable elements to size variations between four teleost genomes. Mob DNA. 7(1):1–16. doi:10.1186/S13100-016-0059-7/FIGURES/8.

George DG, Edwards RW. 1976. The Effect of Wind on the Distribution of Chlorophyll A and Crustacean Plankton in a Shallow Eutrophic Reservoir. J Appl Ecol. 13(3):667. doi:10.2307/2402246.

Gold Z, Sprague J, Kushner DJ, Marin EZ, Barber PH. 2021. eDNA metabarcoding as a biomonitoring tool for marine protected areas. PLoS One. 16(2):e0238557. doi:10.1371/JOURNAL.PONE.0238557.

Goldberg CS, Turner CR, Deiner K, Klymus KE, Thomsen PF, Murphy MA, Spear SF, McKee A, Oyler-McCance SJ, Cornman RS, et al. 2016. Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods Ecol Evol. 7(11):1299–1307. doi:10.1111/2041-210X.12595.

Goricki Š, Stankovic D, Snoj A, Kuntner M, Jeffery WR, Trontelj P, Pavic M, Grizelj Z, Naparus-Aljancic M, Aljancic G. 2017. Environmental DNA in subterranean biology: range extension and taxonomic implications for Proteus. Scientific Reports 2017 7:1. 7(1):1–11. doi:10.1038/srep45054.

Grooten M, Almond REA, editors. 2018. Living Planet Report - 2018: Aiming Higher. WWF, Gland, Switzerland.

Hamada M, Kido Y, Himberg M, Reist JD, Ying C, Hasegawa M, Okada N. 1997. A newly isolated family of short interspersed repetitive elements (SINEs) in coregonid fishes (whitefish) with sequences that are almost identical to those of the *Sma*I family of repeats: Possible evidence for the horizontal transfer of SINEs. Genetics. 146(1):355–367. doi:10.1093/genetics/146.1.355.

Hasselman DJ, Edge TA, Bradford RG. 2009. Discrimination of the Endangered Atlantic Whitefish from Lake Whitefish and Round Whitefish by Use of External Characters. N Am J Fish Manag. 29(4):1046–1057. doi:10.1577/M08-217.1.

Hoban S, Bruford MW, Funk WC, Galbusera P, Griffith MP, Grueber CE, Heuertz M, Hunter ME, Hvilsom C, Stroil BK, et al. 2021. Global Commitments to Conserving and Monitoring Genetic Diversity Are Now Necessary and Feasible. Bioscience. 71(9):964–976. doi:10.1093/BIOSCI/BIAB054.

Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M. 2008. Ecological consequences of genetic diversity. Ecol Lett. 11(6):609–623. doi:10.1111/J.1461-0248.2008.01179.X.

Jensen MR, Sigsgaard ER, Liu S, Manica A, Bach SS, Hansen MM, Møller Petite Rivière, Thomsen PF. 2020. Genome-scale target capture of mitochondrial and nuclear environmental DNA from water samples. Mol Ecol Resour. 21(3):690-702. doi:10.1111/1755-0998.13293.

Jerde CL, Mahon AR, Chadderton WL, Lodge DM. 2011. "Sight-unseen" detection of rare aquatic species using environmental DNA. Conserv Lett. 4(2):150–157. doi:10.1111/j.1755-263X.2010.00158.x.

Jo T, Murakami H, Yamamoto S, Masuda R, Minamoto T. 2019. Effect of water temperature and fish biomass on environmental DNA shedding, degradation, and size distribution. Ecol Evol. 9(3):1135–1146. doi:10.1002/ECE3.4802.

Kazazian HH. 2004. Mobile Elements: Drivers of Genome Evolution. Science. 303(5664):1626–1632. doi:10.1126/SCIENCE.1089670.

Kido Y, Himberg M, Takasaki N, Okada N. 1994. Amplification of Distinct Subfamilies of Short Interspersed Elements During Evolution of the Salmonidae. J Mol Biol. 241(5):633–644. doi:10.1006/JMBI.1994.1540.

Klobucar SL, Rodgers TW, Budy P. 2017. At the forefront: Evidence of the applicability of using environmental DNA to quantify the abundance of fish populations in natural lentic waters with additional sampling considerations. Can J Fish Aquat Sci. 74(12):2030–2034. doi:10.1139/cjfas-2017-0114.

Klymus KE, Merkes CM, Allison MJ, Goldberg CS, Helbing CC, Hunter ME, Jackson CA, Lance RF, Mangan AM, Monroe EM, et al. 2020. Reporting the limits of detection and quantification for environmental DNA assays. Environ DNA. 2(3):271–282. doi:10.1002/edn3.29.

Klymus KE, Richter CA, Chapman DC, Paukert C. 2015. Quantification of eDNA shedding rates from invasive bighead carp *Hypophthalmichthys nobilis* and silver carp *Hypophthalmichthys molitrix*. Biol Conserv. 183:77–84. doi:10.1016/J.BIOCON.2014.11.020.

Kothe M, Seidenberg V, Hummel S, Piskurek O. 2016. *Alu* SINE analyses of 3,000-year-old human skeletal remains: A pilot study. Mob DNA. 7(1):1–9. doi:10.1186/S13100-016-0063-Y/TABLES/3.

Kramerov DA, Vassetzky NS. 2005. Short Retroposons in Eukaryotic Genomes. Int Rev Cytol. 247:165–221. doi:10.1016/S0074-7696(05)47004-7.

Kramerov DA, Vassetzky NS. 2011. Origin and evolution of SINEs in eukaryotic genomes. Heredity. 107(6):487–495. doi:10.1038/hdy.2011.43.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, Fitzhugh W, et al. 2001. Initial sequencing and analysis of the human genome. Nature. 409(6822):860–921. doi:10.1038/35057062.

Lerch A, Koepfli C, Hofmann NE, Messerli C, Wilcox S, Kattenberg JH, Betuela I, O'Connor L, Mueller I, Felger I. 2017. Development of amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections. BMC Genomics. 18(1). doi:10.1186/S12864-017-4260-Y.

Li J, Lawson Handley LJ, Read DS, Hänfling B. 2018. The effect of filtration method on the efficiency of environmental DNA capture and quantification via metabarcoding. Mol Ecol Resour. 18(5):1102–1114. doi:10.1111/1755-0998.12899.

Liang JW, Coyle HM. 2020. A short interspersed nuclear element-based quantitative PCR assay for simultaneous human and dog DNA detection and quantification. Biotechniques. 70(3):175–180. doi:10.2144/BTN-2020-0144.

Lodge DM, Williams S, Macisaac HJ, Hayes KR, Leung B, Reichard S, Mack RN, Moyle PB, Smith M, Andow DA, et al. 2006. ESA report biological invasions: recommendations for u.s. policy and management. Ecol Appl. 16(6):2035–2054. doi:10.1890/1051-0761(2006)016.

Loeza-Quintana T, Abbott CL, Heath DD, Bernatchez L, Hanner RH. 2020. Pathway to Increase Standards and Competency of eDNA Surveys (PISCeS)—Advancing collaboration and standardization efforts in the field of eDNA. Environ DNA. 2(3):255–260. doi:10.1002/EDN3.112.

Madden T. 2002. The BLAST Sequence Analysis Tool. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Chapter 16. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21097/

Marshall NT, Stepien CA. 2019. Invasion genetics from eDNA and thousands of larvae: A targeted metabarcoding assay that distinguishes species and population variation of zebra and quagga mussels. Ecol Evol. 9(6):3515–3538. doi:10.1002/ECE3.4985.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17(1):10–12.

Maruyama A, Nakamura K, Yamanaka H, Kondoh M, Minamoto T. 2014. The Release Rate of Environmental DNA from Juvenile and Adult Fish. Stöck M, editor. PLoS One. 9(12):e114639. doi:10.1371/journal.pone.0114639.

Mathon L, Valentini A, Guérin PE, Normandeau E, Noel C, Lionnet C, Boulanger E, Thuiller W, Bernatchez L, Mouillot D, et al. 2021. Benchmarking bioinformatic tools for fast and accurate eDNA metabarcoding species identification. Mol Ecol Resour. 21(7):2565–2579. doi:10.1111/1755-0998.13430.

Matveev V, Okada N. 2009. Retroposons of salmonoid fishes (Actinopterygii: Salmonoidei) and their evolution. Gene. 434(1–2):16–28. doi:10.1016/J.GENE.2008.04.022.

Mérot C, Stenløkk KSR, Venney C, Laporte M, Moser M, Normandeau E, Árnyasi M, Kent M, Rougeux C, Flynn JM, et al. 2023. Genome assembly, structural variants, and genetic differentiation between lake whitefish young species pairs (*Coregonus* sp.) with long and short reads. Mol Ecol. 32(6):1458–1477. doi:10.1111/MEC.16468.

Minamoto T, Uchii K, Takahara T, Kitayoshi T, Tsuji S, Yamanaka H, Doi H. 2017. Nuclear internal transcribed spacer-1 as a sensitive genetic marker for environmental DNA studies in common carp *Cyprinus carpio*. Mol Ecol Resour. 17(2):324–333. doi:10.1111/1755-0998.12586.

Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T, Yamamoto S, Yamanaka H, Araki H, et al. 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. R Soc Open Sci. 2(7). doi:10.1098/RSOS.150088.

Murray K. 2005. Population genetic assessment of the endangered Atlantic Whitefish, *Coregonus huntsmani*, and the Lake Whitefish, *C. clupeaformis*, in Atlantic Canada. M.Sc. thesis. Dalhousie University, Halifax, Nova Scotia. 79pp.

Nevers MB, Byappanahalli MN, Morris CC, Shively D, Przybyla-Kelly K, Spoljaric AM, Dickey J, Roseman EF. 2018. Environmental DNA (eDNA): A tool for quantifying the abundant but elusive round goby (*Neogobius melanostomus*). PLoS One. 13(1):e0191720. doi:10.1371/JOURNAL.PONE.0191720.

Nolan KP, Loeza-Quintana T, Little HA, McLeod J, Ranger B, Borque DA, Hanner RH. 2023. Detection of brook trout in spatiotemporally separate locations using validated eDNA technology. J Environ Stud Sci. 13(1):66–82. doi:10.1007/S13412-022-00800-X/FIGURES/9.

Parsons KM, Everett M, Dahlheim M, Park L. 2018. Water, water everywhere: Environmental DNA can unlock population structure in elusive marine species. R Soc Open Sci. 5(8). doi:10.1098/rsos.180537.

Pauvert C, Buée M, Laval V, Edel-Hermann V, Fauchery L, Gautier A, Lesur I, Vallance J, Vacher C. 2019. Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. Fungal Ecol. 41:23–33. doi:10.1016/J.FUNECO.2019.03.005.

Pinfield R, Dillane E, Runge AKW, Evans A, Mirimin L, Niemann J, Reed TE, Reid DG, Rogan E, Samarra FIP, et al. 2019. False-negative detections from environmental DNA collected in the presence of large numbers of killer whales (*Orcinus orca*). Environ DNA. 1(4):316–328. doi:10.1002/EDN3.32.

R Core Team. 2016. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing

Rishishwar L, Tellez Villa CE, Jordan IK. 2015. Transposable element polymorphisms recapitulate human evolution. Mob DNA. 6(1):1–13. doi:10.1186/S13100-015-0052-6/FIGURES/5.

Robin ED, Wong R. 1988. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol. 136(3):507–513. doi:10.1002/JCP.1041360316.

Rodgers EM. 2021. Adding climate change to the mix: Responses of aquatic ectotherms to the combined effects of eutrophication and warming. Biol Lett. 17(10). doi:10.1098/rsbl.2021.0442.

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 4(10). doi:10.7717/PEERJ.2584.

Rojahn J, Pearce L, Gleeson DM, Duncan RP, Gilligan DM, Bylemans J. 2021. The value of quantitative environmental DNA analyses for the management of invasive and endangered native fish. Freshw Biol. 00:1–11. doi:10.1111/FWB.13779.

Sakata MK, Watanabe T, Maki N, Ikeda K, Kosuge T, Okada H, Yamanaka H, Sado T, Miya M, Minamoto T. 2021. Determining an effective sampling method for eDNA metabarcoding: a case study for fish biodiversity monitoring in a small, natural river. Limnology (Tokyo). 22(2):221–235. doi:10.1007/s10201-020-00645-9.

Sambrook J, Fritschi EF, Maniatis T. 1989. Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press, New York.

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, et al. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science. 274(5288):765–768. doi:10.1126/SCIENCE.274.5288.765.

Sansom BJ, Sassoubre LM. 2017. Environmental DNA (eDNA) Shedding and Decay Rates to Model Freshwater Mussel eDNA Transport in a River. Environ Sci Technol. 51(24):14244–14253. doi: 10.1021/acs.est.7b05199.

Sassoubre LM, Yamahara KM, Gardner LD, Block BA, Boehm AB. 2016. Quantification of Environmental DNA (eDNA) Shedding and Decay Rates for Three Marine Fish. Environ Sci Technol. 50(19):10456–10464. doi: 10.1021/acs.est.6b03114.

Scott WB. 1987. A new name for the Atlantic whitefish: *Coregonus huntsmani* to replace *Coregonus canadensis*. Can J Zool. 65

Sigsgaard EE, Carl H, Møller PR, Thomsen PF. 2015. Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. Biol Conserv. 183:46–52. doi: 10.1016/j.biocon.2014.11.023

Sigsgaard EE, Jensen MR, Winkelmann IE, Møller PR, Hansen MM, Thomsen PF. 2020. Population-level inferences from environmental DNA—Current status and future perspectives. Evol Appl. 13(2):245. doi:10.1111/EVA.12882.

Sigsgaard EE, Nielsen IB, Bach S, Lorenzen E, Robinson D, Knudsen SW, Pedersen M, Jaidah M Al, Orlando L, Willerslev E, et al. 2016. Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. Nat Ecol Evol. 1(1):4. doi: 10.1038/s41559-016-0004.

Smith DP, Peay KG. 2014. Sequence Depth, Not PCR Replication, Improves Ecological Inference from Next Generation DNA Sequencing. PLoS One. 9(2):e90234. doi:10.1371/JOURNAL.PONE.0090234.

Sotero-Caio CG, Platt RN, Suh A, Ray DA. 2017. Evolution and Diversity of Transposable Elements in Vertebrate Genomes. Genome Biol Evol. 9(1):161–177. doi:10.1093/GBE/EVW264.

Stapley J, Santure AW, Dennis SR. 2015. Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. Mol Ecol. 24(9):2241–2252. doi:10.1111/MEC.13089.

Stat M, Huggett MJ, Bernasconi R, Dibattista JD, Berry TE, Newman SJ, Harvey ES, Bunce M. 2017. Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. Scientific Reports 2017 7:1. 7(1):1–11. doi:10.1038/s41598-017-12501-5.

Stepien CA, Snyder MR, Elz AE. 2019. Invasion genetics of the silver carp *Hypophthalmichthys molitrix* across North America: Differentiation of fronts, introgression, and eDNA metabarcode detection. PLoS One. 14(3):e0203012. doi:10.1371/JOURNAL.PONE.0203012.

Svec D, Tichopad A, Novosadova V, Pfaffl MW, Kubista M. 2015. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. Biomol Detect Quantif. 3:9. doi:10.1016/J.BDQ.2015.01.005.

Takahara T, Minamoto T, Yamanaka H, Doi H, Kawabata Z. 2012. Estimation of fish biomass using environmental DNA. PLoS One. 7(4). doi:10.1371/journal.pone.0035868.

Takahashi M, Saccò M, Kestel JH, Nester G, Campbell MA, van der Heyde M, Heydenrych MJ, Juszkiewicz DJ, Nevill P, Dawkins KL, et al. 2023. Aquatic environmental DNA: A review of the macro-organismal biomonitoring revolution. Sci Total Environ. 873:162322. doi:10.1016/J.SCITOTENV.2023.162322.

Themelis DE, Bradford RG, Leblanc PH, O'neil SF, Breen AP, Longue P, Nodding SB. 2014. Monitoring activities in support of endangered Atlantic Whitefish (*Coregonus huntsmani*) recovery efforts in the Petite Rivière lakes in 2013. Can. Manuscr. Rep. Fish. Aquat, Sci. 3031. v + 94 p.

Tickner D, Opperman JJ, Abell R, Acreman M, Arthington AH, Bunn SE, Cooke SJ, Dalton J, Darwall W, Edwards G, et al. 2020. Bending the Curve of Global Freshwater Biodiversity Loss: An Emergency Recovery Plan. Bioscience. 70(4):330–342. doi:10.1093/BIOSCI/BIAA002.

Tillotson MD, Kelly RP, Duda JJ, Hoy M, Kralj J, Quinn TP. 2018. Concentrations of environmental DNA (eDNA) reflect spawning salmon abundance at fine spatial and temporal scales. Biol Conserv. 220:1–11. doi:10.1016/j.biocon.2018.01.030.

Turon X, Antich A, Palacín C, Præbel K, Wangensteen OS. 2020. From metabarcoding to metaphylogeography: separating the wheat from the chaff. Ecol Appl. 30(2):e02036. doi:10.1002/EAP.2036.

Uchii K, Doi H, Minamoto T. 2016. A novel environmental DNA approach to quantify the cryptic invasion of non-native genotypes. Mol Ecol Resour. 16(2):415–422. doi:10.1111/1755-0998.12460.

Wells JN, Feschotte C. 2020. A Field Guide to Eukaryotic Transposable Elements. Annu Rev Genet. 54:539-561. doi:10.1146/annurev-genet-040620-022145

Weltz K, Lyle JM, Ovenden J, Morgan JAT, Moreno DA, Semmens JM. 2017. Application of environmental DNA to detect an endangered marine skate species in the wild. PLoS One. 12(6):e0178124. doi:10.1371/JOURNAL.PONE.0178124.

Weusten J, Herbergs J. 2012. A stochastic model of the processes in PCR based amplification of STR DNA in forensic applications. Forensic Sci Int Genet. 6(1):17–25. doi:10.1016/J.FSIGEN.2011.01.003.

Whitelaw J, Manríquez-Hernández J, Duston J, O'Neil SF, Bradford RG. 2015. Atlantic Whitefish (*Coregonus huntsmani*) culture handbook. Can. Manuscr. Rep. Fish. Aquat. Sci. 3074: vii + 55 p.

Williams-Subiza EA, Epele LB. 2021. Drivers of biodiversity loss in freshwater environments: A bibliometric analysis of the recent literature. Aquat Conserv. 31(9):2469–2480. doi:10.1002/AQC.3627.

Wood ZT, Erdman BF, York G, Trial JG, Kinnison MT. 2020. Experimental assessment of optimal lotic eDNA sampling and assay multiplexing for a critically endangered fish. Environ DNA. 2(4):407–417. doi:10.1002/edn3.64.

Xia Z, Zhan A, Johansson ML, Deroy E, Haffner GD, Macisaac HJ. 2021. Screening marker sensitivity: Optimizing eDNA-based rare species detection. Divers Distrib. 27(10):1981-1988. doi:10.1111/ddi.13262.

Xie R, Zhao G, Yang J, Wang Zhihao, Xu Y, Zhang X, Wang Zijian. 2021. eDNA metabarcoding revealed differential structures of aquatic communities in a dynamic freshwater ecosystem shaped by habitat heterogeneity. Environ Res. 201:111602. doi:10.1016/J.ENVRES.2021.111602.

Young MK, Isaak DJ, Nagel D, Horan DL, Carim KJ, Franklin TW, Zeller VA, Roper B, Schwartz MK. 2022. Broad-scale eDNA sampling for describing aquatic species distributions in running waters: Pacific lamprey *Entosphenus tridentatus* in the upper Snake River, USA. J Fish Biol. 101(5):1312–1325. doi:10.1111/JFB.15202.

Zanovello L, Girardi M, Marchesini A, Galla G, Casari S, Micheletti D, Endrizzi S, Fedrigotti C, Pedrini P, Bertorelle G, et al. 2023. A validated protocol for eDNA-based monitoring of within-species genetic diversity in a pond-breeding amphibian. Sci Rep. 13(1):1–10. doi:10.1038/s41598-023-31410-4.

Zhang C, Wang L, Dou L, Yue B, Xing J, Li J. 2023. Transposable Elements Shape the Genome Diversity and the Evolution of Noctuidae Species. Genes (Basel). 14(6):1244. doi:10.3390/GENES14061244/S1.

Zhang S, Lu Q, Wang Y, Wang X, Zhao J, Yao M. 2020. Assessment of fish communities using environmental DNA: Effect of spatial sampling design in lentic systems of different sizes. Mol Ecol Resour. 20(1):242–255. doi:10.1111/1755-0998.13105.

Zhang T, Fang HHP. 2006. Applications of real-time polymerase chain reaction for quantification of microorganisms in environmental samples. Appl Microbiol Biotechnol. 70(3):281–289. doi:10.1007/S00253-006-0333-6/FIGURES/3.

Appendix A – qPCR assays

Table A1 Hydrolysis assay primers and probes targeting the ND4 subunit and *Sma*I-corII SINE in Atlantic Whitefish.

Marker	ND4	*Sma*I-corII
DNA type	mtDNA	nuDNA
Forward Primer (5'-3')	ACTGACCTGTTGGCTGCTTC	TAGCTCAGCTGGTAGAGCAC
Reverse Primer (5'-3')	GTTCTGTTTGGTTGCCCCATC	TGCCATTTAGCAGACGCTTTT
Probe (5'-3')	ATCCTCGCAAGTCAAAACCACATCAACC	AACGCCAAGGTAGTGGGTTCGATC
Amplicon length	224 bp	131 bp

Appendix B – 2023 qPCR analysis of net pen-associated samples

Following optimization of the gBlock standards, the net pen-associated samples were reanalyzed via qPCR for each assay in order to resolve concentration (Figure B1). For *Sma*I-corII, the efficiency, error, and slope were 2.03, 0.06, and -3.26, respectively. For ND4 the efficiency, error, and slope were 1.99, 0.07, and -3.35, respectively. This subsequent qPCR occurred approximately one year following the initial qPCR. One PCR replicate of the field blank amplified with a CP of 40.52 with the *Sma*I-corII assay while no blanks amplified with ND4. Observed CPs were higher than in the first qPCR (Figure B2). Welch's t-test found significant differences in the mean CP of the 0 m (t(13.24)= -6.15, p-value = 3.218e-05) and 20 m (t(12.50) = -4.11, p-value = 0.001) samples for the *Sma*I-corII assay (Table B1). No significant difference was observed in the 0 m sample for the ND4 assay between the 2022 and 2023 qPCRs (t(15.87) = -0.10, p-value = 0.925). As the ND4 2023 qPCR only had amplification in the 0 m sample, no statistical comparisons to 2022 could be made for the other distances.

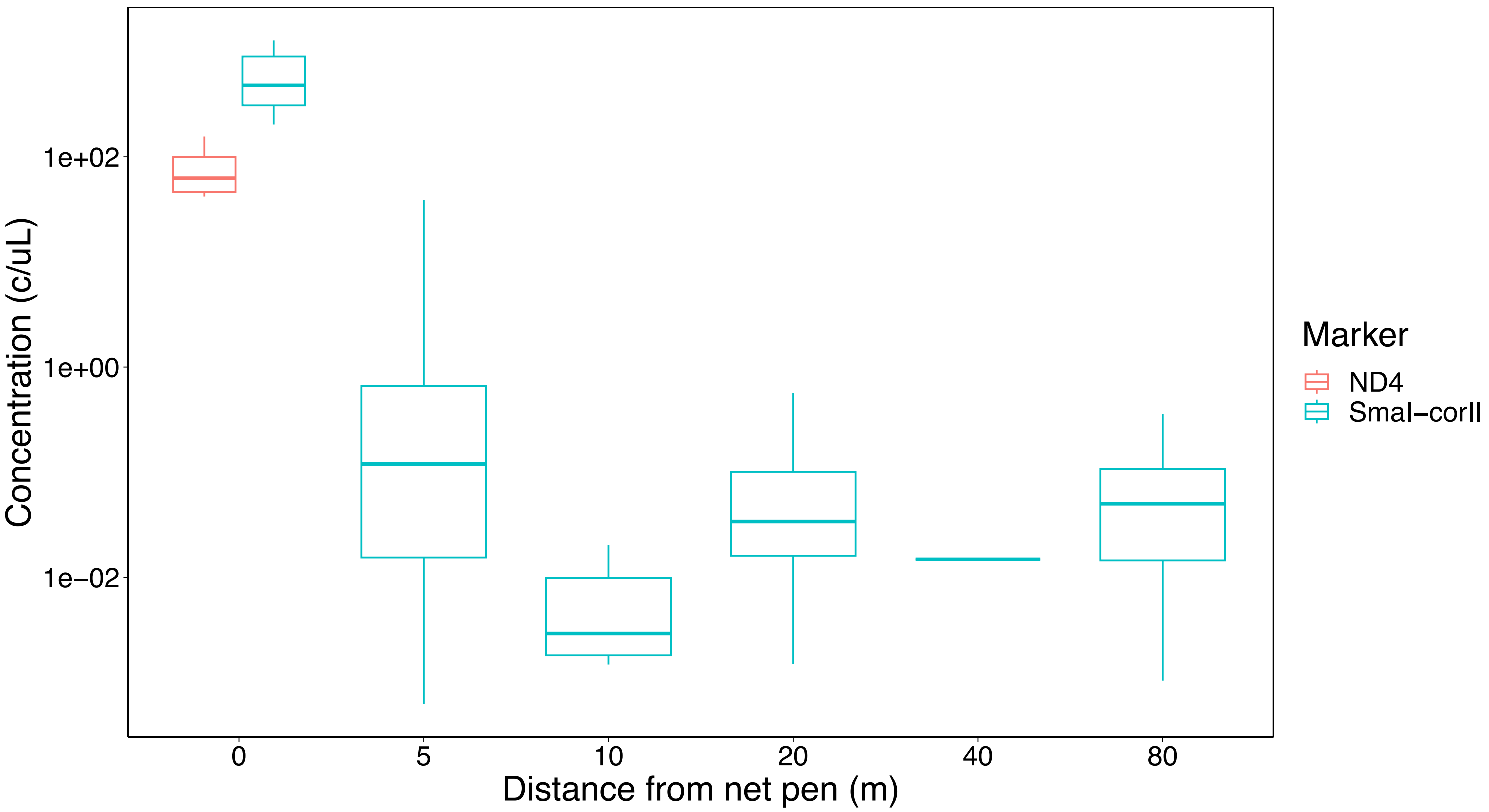

Figure B1 Concentration (c/µL, log-10 scale) of eDNA samples amplified with *Sma*I-corII and ND4 assays taken up to 80 m away from a net pen housing 150 ~3g juvenile Atlantic Whitefish in Milipsigate Lake, NS one year after sample collection.

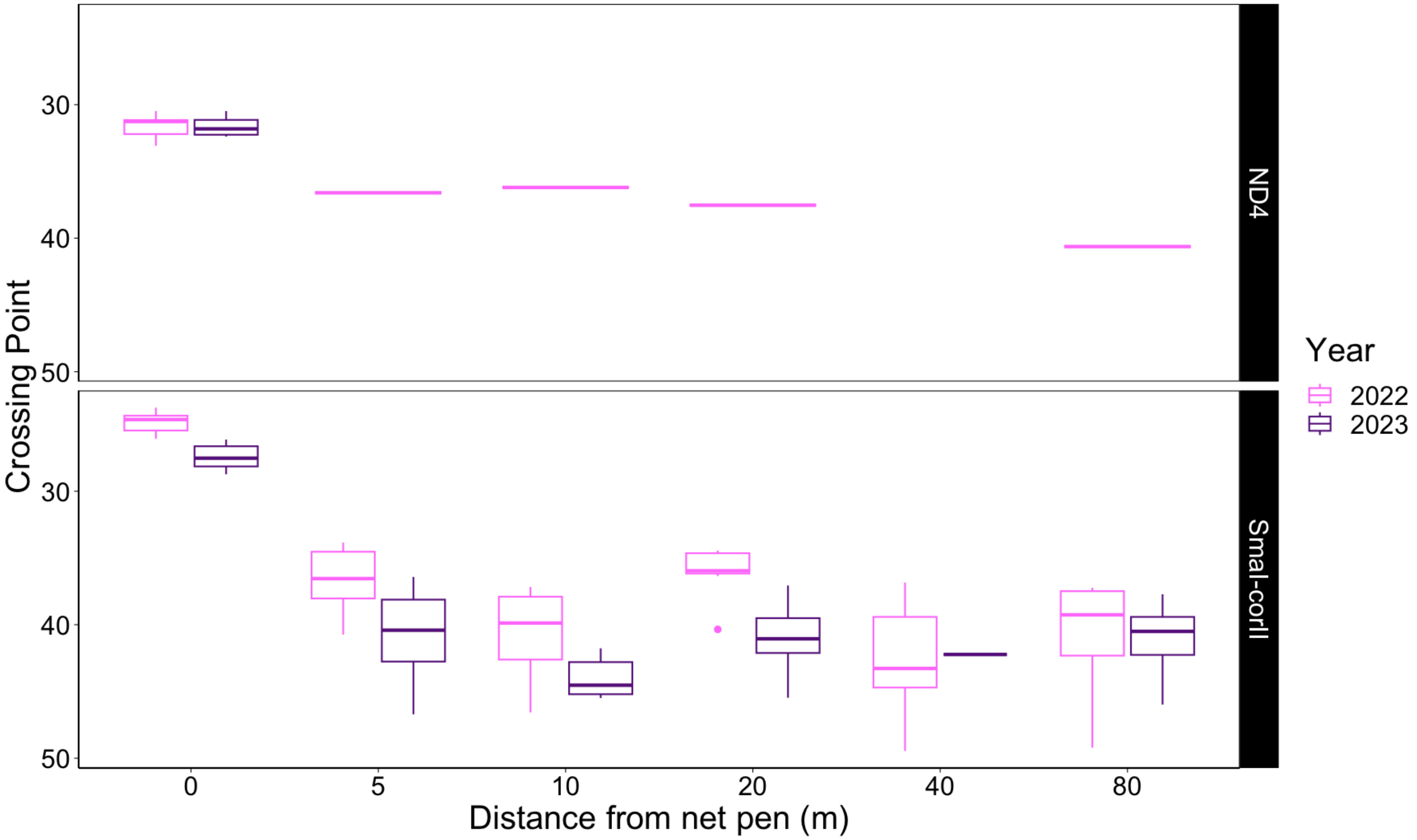

Figure B2 Comparison of CPs from qPCRs performed in 2022 and 2023 on samples taken up to 80 m away from a net pen housing 150 ~3g juvenile Atlantic Whitefish for *Sma*I-corII and ND4 assays.

Table B1 Welch's t-test results of mean CP from 2022 and 2023 qPCRs of net pen-associated samples for *Sma*I-corII and ND4 assays. Significant differences in mean CP between the years are represented by bolded values.

Distance (m)	ND4	***Sma*I-corII**
0	t(15.87) = -0.10 , p-value = 0.925	**t(13.24) = -6.15, p-value = 3.218e-05**
5	-	t(6.68) = -2.00, p-value = 0.088
10	-	t(6.55) = -2.04, p-value = 0.084
20	-	**t(12.47) = -4.11, p-value = 0.001**
40	-	t(1.02) = 1.03, p-value = 0.490
80	-	t(7.93) = -0.17, p-value = 0.870

.

Appendix C – 2023 qPCR analysis of density trial samples

As with the net pen-associated samples, following gBlock optimization the density trial samples were subjected to a second qPCR roughly one year after the first. For the *Sma*I-corII assay, the efficiency, error, and slope of the qPCR were 1.66, 0.07, and -4.54, respectively. The efficiency, error, and slope of the ND4 assay were 1.96, 0.06, and -3.42, respectively. No amplification was observed in the control tank which contained no Atlantic Whitefish for either assay. For each assay, the concentrations were resolved (Figure C1); however, concentrations following the *Sma*I-corII assay were extrapolated from the standard curve as they crossed the background noise threshold before the highest standard. Before re-running the *Sma*I-corII qPCR with higher standards and generating a more reliable standard curve (Zhang and Fang 2006), the corresponding CPs were compared to those determined in the initial 2022 qPCR. Observed 2023 CPs were higher than those from the 2022 qPCR, indicating possible sample degradation and so samples were not reanalyzed with higher standards (Figure C2). Higher CPs in the 2023 qPCR compared to the 2022 qPCR were observed in the ND4 assay as well. In comparing the mean CP following the ND4 assay of each tank from the 2022 and 2023 qPCRs, differences were significant for all tanks (Table C1). Significant differences in mean CP following the *Sma*I-corII assay were observed for tank 2 (t(8) = -7.40, p = 7.63e-05, Welch's t-test) and tank 3 (t(10.38) = -8.24, p = 7.17e-06, Welch's t-test). Differences in mean CP between tanks 4 and 5 were not significant.

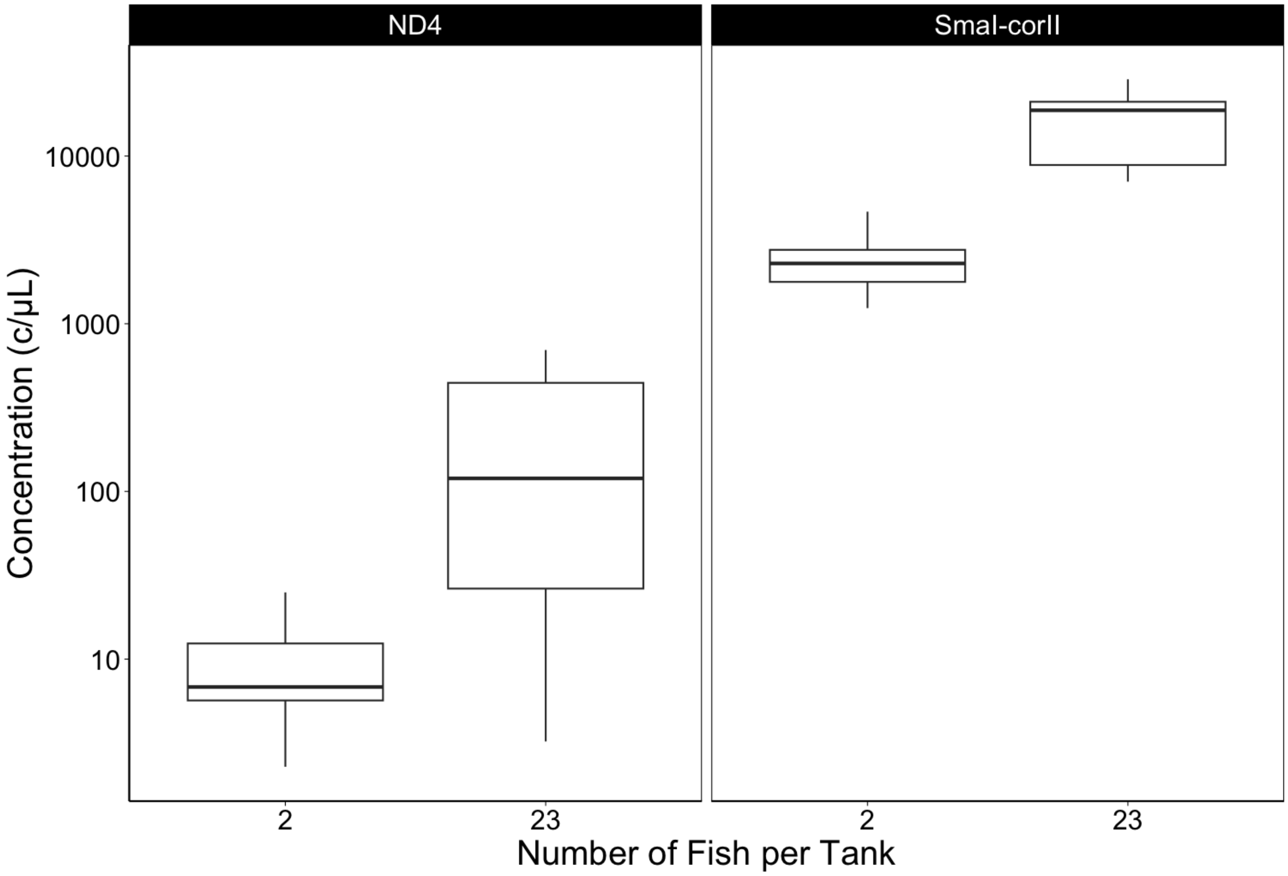

Figure C1 Concentration (c/µL, log10-scale) of samples taken from tanks holding 2 or 23 ~3g juvenile Atlantic Whitefish for *Sma*I-corII and ND4 assays.

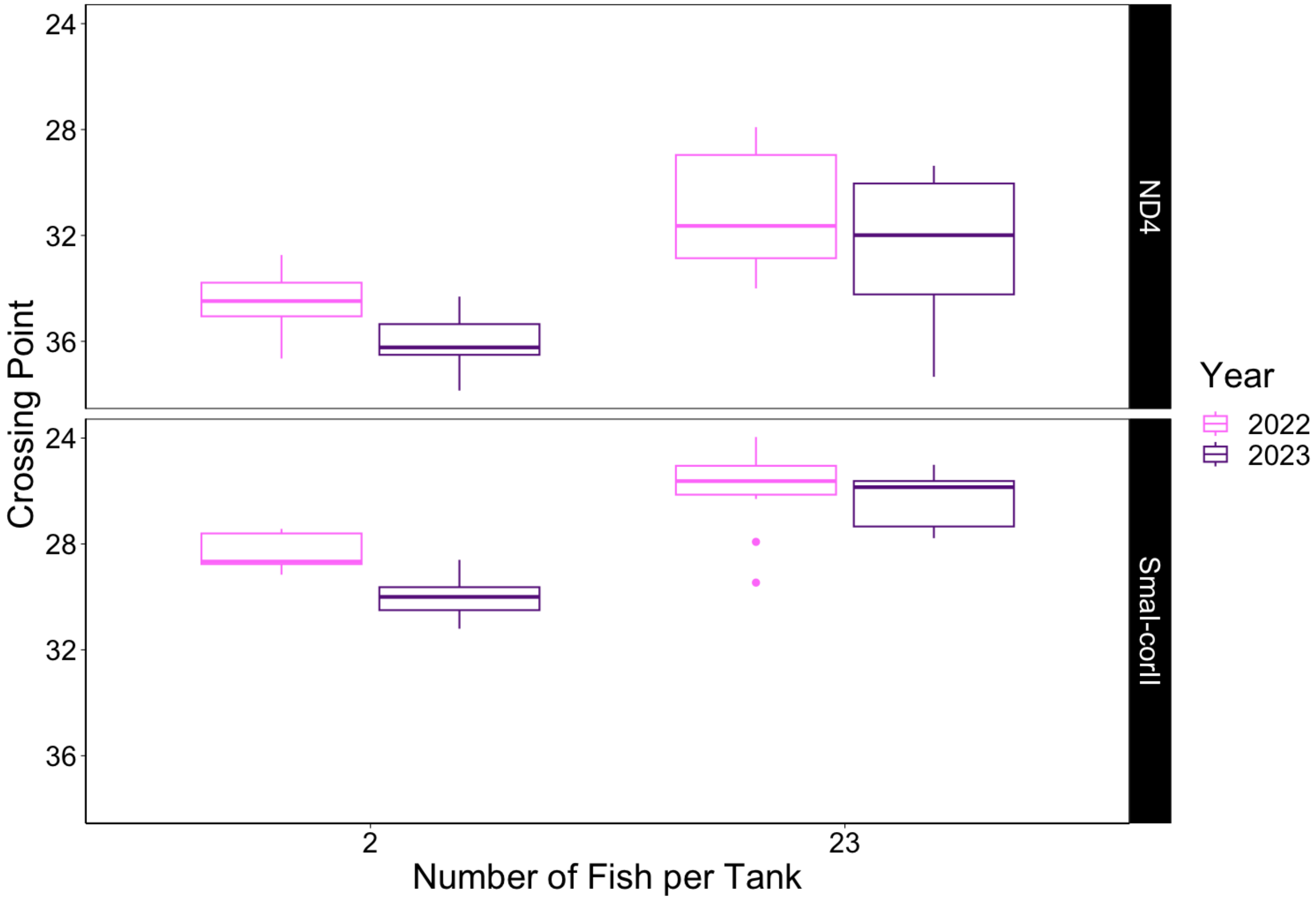

Figure C2 Comparison of crossing points from qPCRs performed in 2022 and 2023 on samples taken from tanks housing 2 or 23 juvenile Atlantic Whitefish amplified with ND4 and *Sma*I-corII assays.

Table C1 Welch's t-test results of mean CP from 2022 and 2023 qPCRs of density trial samples (0.45 µM pore size) for *Sma*I-corII and ND4 assays. Significant differences in mean CP between the years are represented by bolded values.

Tank	Number of fish	ND4	*Sma*I-corII
1	0	-	-
2	2	**t(14.59) = -3.65, p = 0.002**	**t(8) = -7.40, p = 7.63e-05**
3	2	**t(12.89) = -2.57 , p = 0.023**	**t(10.38) = -8.24, p = 7.17e-06**
4	23	**t(13.29) = -3.09, p = 0.008**	t(14.24) = -1.42, p = 0.177
5	23	**t(12.67) = -3.51 , p = 0.004**	t(9.55) = -0.56, p = 0.590

Appendix D – 2023 filter pore size comparison

eDNA yield obtained with the 5 µM prefilter and 0.45 µM filter was also compared via qPCR for the *Sma*I-corII assay in 2023. The efficiency, error, and slope of the 5 µM prefilter qPCR were 1.74, 0.06, and -4.16, respectively, which are slightly outside of the acceptable range for a standard curve. As this analysis was performed to compare CP values on samples with observed degradation, the curve was not generated and concentrations of the samples were not obtained. Amplification was observed in 2/9 replicates taken from within the control tank (CP = 43.36 ± 5.62) with the 5 µM filter and no amplification was observed in the field blank. No amplification was observed in the control tank or field blank for the 0.45 µM filter. CPs were used to compare eDNA yield as concentrations were extrapolated from the standard curve and not reliably resolved for both filter pore sizes. The 23-fish tanks had a significant difference in mean CP for the 5 µM prefilters (t(12.97) = -0.09, p = 0.930, Welch's t-test) and no significant difference between the 2-fish tanks (t(14.88) = -4.20, p = 0.0008, Welch's t-test), a pattern opposite to what was observed with the 0.45 µM filters. For the 2-fish samples, the 5 µM prefilter had a lower CP than the 0.45 µM filter, indicating higher eDNA yield (Figure D1). For the 23-fish samples, CPs were similar for each filter pore size, however greater variance was observed across the 5 µM prefilters. On the 0.45 µM filter, the mean CP of the 2-fish tanks was 5.29 cycles greater than the mean CP of the 5 µM prefilter (0.45 µM = 29.90 ±0.78, 5 µM = 24.61 ±1.86), which corresponds to 39 times more eDNA on the prefilter than the filter. The difference in mean CP between the pore sizes at the 2-fish density was significant (t(22.81) = -11.13, p = 1.083e-10, Welch's t-test). No significant difference was observed in mean CP of the 23-fish samples between the two pore sizes (t(18.66) = -0.65, p = 0.526, Welch's t-test), with the 5 µM filters having approximately 1.6 times more eDNA than the 0.45 filter (0.45 µM = 26.29

±0.95, 5 µM = 25.63 ± 4.29). Assessing CPs on a per tank basis, the 5 µM filter had lower CPs for all tanks except tank 5 and significant differences in CP were observed across the filter pore sizes for each tank (Table D1).

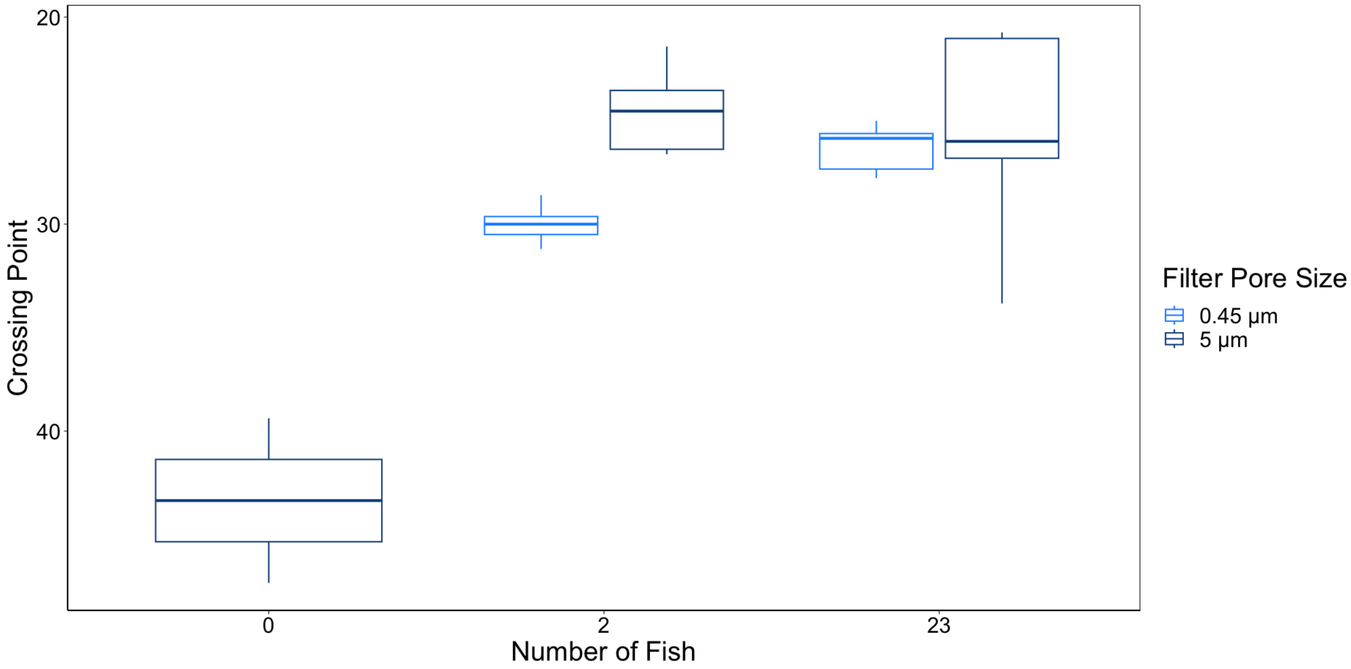

Figure D1 Comparison of mean CP for density trial samples filtered onto a 5 µM prefilter or 0.45 µM filter.

Table D1 Average crossing point (CP), number of technical replicates with detection, and Welch's t-test results for density trial samples filtered onto a 5 μM prefilter or 0.45 μM filter. Significant differences in CP between the pore sizes are represented by bolded values.

		0.45 μM		**5 μM**		
Tank	Number of fish	CP (μ ± SE)	Replicates with amplification	CP (μ ± SE)	Replicates with amplification	Difference in CP between filter pore sizes
Blank	-	-	0/3	-	0/3	-
1	0	-	0/9	43.36 (±5.62)	2/9	-
2	2	29.51 (±0.85)	9/9	24.57 (±2.33)	9/9	**t(10.09) = -5.98, p = 0.0001**
3	2	30.28 (±0.50)	9/9	24.65 (±1.37)	9/9	**t(10.05) = -11.53, p = 4.059e-07**
4	23	26.92 (±0.94)	9/9	22.61 (±2.59)	9/9	**t(10.05) = -4.70, p = 0.0008**
5	23	25.65 (±0.36)	9/9	28.64 (±3.44)	9/9	**t(8.17) = 2.59, p = 0.032**

Appendix E – Overview of a generalized bioinformatic pipeline to process eDNA metabarcoding NGS data

Following DNA sequencing, returned sequences must be assessed via bioinformatic pipelines (hereafter, ‘pipelines’), suites of connected algorithms executed in a specific order to analyze next generation sequencing (NGS) data. Many pipelines are available, and the general mechanisms of sequence manipulation are similar, though the choice of program and order in which programs are executed may vary (Mathon et al. 2021).

Generally, samples are firstly demultiplexed where input samples, identified via unique indices, are split into individual fastq files. Some sequencing machines are capable of performing this step. Secondly, sequences of interest are identified by their primers, which are then removed (Figure E1). If paired-end sequencing (recommended) was used, forward and reverse reads are merged based on sequence quality (which generally declines along the length of the read) and read overlap to assemble the complete amplicon sequence. Reads then undergo quality filtering to remove any sequences which contain sequencing errors, as determined by the probability of a base calling error (“phred scores”). Common filtering methods include truncating or excluding reads based on phred scores, sequence length, and ambiguous bases (Bokulich et al. 2013). Reads may also be trimmed to a set length. The dataset is then simplified via dereplication, where unique sequences are retained once alongside an abundance metric to reflect how many copies of that exact sequence were present in each sample prior to dereplication. Denoising, a form of error correction, is then often employed.

Most commonly, during denoising, sequences are compared to one another based on abundance using an error model to detect putatively correct sequences to which the putatively incorrect sequences are clustered (Edgar 2016a). Highly abundant sequences form a centroid which all other sequences are compared to based on their dissimilarity. Sequences that are not dissimilar enough to the centroid are inferred to be errors of the centroid and are clustered within it, essentially hiding the erroneous sequences behind the centroid. This process is iterative, where all sequences are compared to the centroid. After the initial comparison, the next highly abundant sequence forms a centroid, and the process repeats until all sequences have been compared to one another. Following denoising, only the centroid sequences remain. Chimeric sequences are then identified and removed. Taxonomy of remaining sequences can then be inferred by comparison to a reference database. As such, the quality and diversity of sequences within the database will have a large effect on obtained results (Blackman et al. 2023).

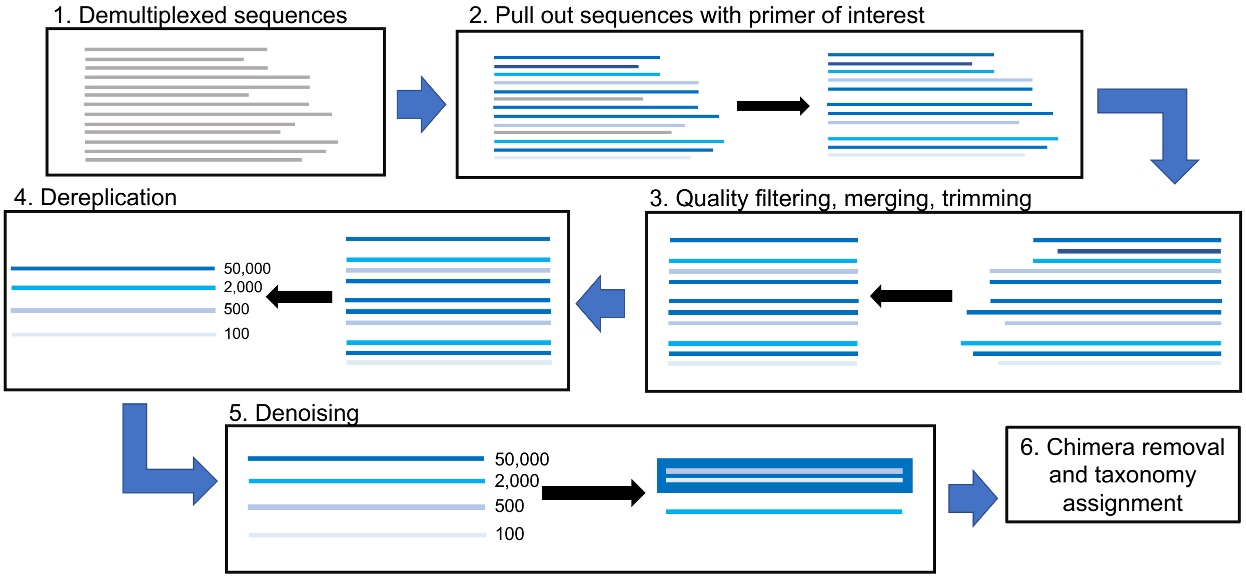
1. Demultiplexed sequences
2. Pull out sequences with primer of interest
3. Quality filtering, merging, trimming
4. Dereplication
50,000
2,000
500
100
5. Denoising
50,000
2,000
500
100
6. Chimera removal and taxonomy assignment

Figure E1 Overview of a generalized bioinformatic pipeline to process eDNA metabarcoding NGS data.

Appendix F – Initial optimization of custom and DADA2 pipelines and final comparison

Following numerous iterations of pipeline optimization for the custom and DADA2 pipelines which focused heavily on the primer removal and denoising stages (Tables F1, F2), *Lobelia* spp. sequences were identified in the demultiplexed input files. The presence of these sequences within the input files indicated that the 2,197,228 single-end reads obtained following sequencing were not reflective of the number of *Sma*I-corII sequences within the file and not a reliable value of primer removal parameterization (Figure F1). A core aspect of the DADA2 pipeline is its error estimation and correction functions which masked the presence of *Lobelia* spp. sequences within the analysis. Turning off the error correction parameter (OMEGA_C) resulted in 526,049 reads across 1,598 ASVs in the final DADA2 output (Figure F2), only 14 of which were consistently detected across PCR replicates (0.88%) totaling 168,389 reads (32%). The large discrepancy between reads retained with and without the error correction parameter indicate *Lobelia* spp. sequences were being corrected to *Sma*I-corII centroids, thereby increasing the depth. Therefore the error correction estimate and correction functions of DADA2 were determined to be unsuitable for *Sma*I-corII variant assessment and the optimization was restarted for the custom pipeline only, noting the presence of *Lobelia* spp. in the input files.

Table F1 Parameter changes tested with the custom pipeline where an asterisks (*) represents the same parameter value as the prior optimization attempt.

	Pipeline step and parameters					
Optimization attempt	**Primer removal**	**Quality filtering**	**Dereplication**	**Denoising**		**Parameterization Description**
	e	**maxEE**	**minlen**	**minsize**	**unoise_alpha**	
1	0.1	1	32	8	2	All default parameters. No taxonomy assignment.
2	0.15	*	*	*	*	Allow MORE mismatches in priming region.
3	0.5	*	*	*	*	Allow MORE mismatches in priming region.
4	*	2	*	*	*	LESS stringent quality filtering.
5	0.4	1	*	*	*	Allow LESS mismatches in priming region.
6	*	*	*	*	1	LESS stringent quality filtering.
7	*	2	*	*	*	MORE stringent denoising.
8	*	*	*	4	2	REDUCE frequency sequences must be observed to be retained.
9	*	*	*	*	1	MORE stringent denoising.
10	*	*	*	*	3	LESS stringent denoising
11	0.1	*	*	*	2	Tried to match DADA2's default parameters by the following during quality filtering: fastq_maxns 0 fastq_truncqual 2 fastq_minlen 50
12	0.4	*	*	*	5	LESS stringent denoising
13	*	*	*	*	10	LESS stringent denoising

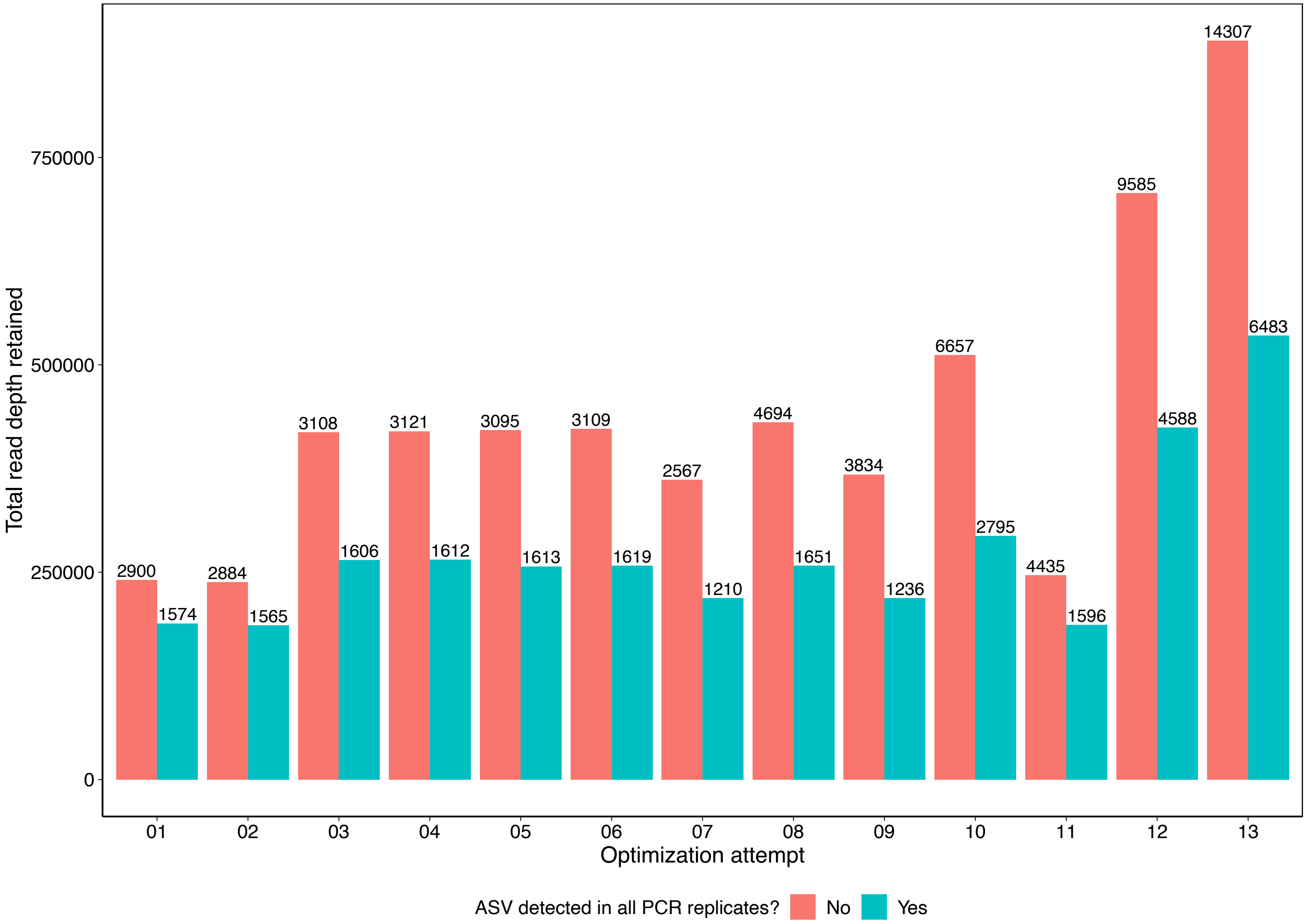

Figure F1 Total read depth retained across custom pipeline optimization attempts. Parameter changes corresponding to the optimization attempt are detailed in Table F1. The value above each bar represents how many ASVs were contained within the retained read depth.

Table F2 Parameter changes tested with the DADA2 pipeline where an asterisks (*) represents the same parameter value as the prior optimization attempt.

Optimization attempt	Pipeline step and parameters									Parameterization Description
	Primer removal		Quality filtering				Denoising			
	e	minimum-length	maxN	maxEE	truncQ	minLen	OMEGA_A	OMEGA_C	BAND_SIZE	
1	0.1	-	0	2	2	50	1e-40	1e-40	16	All default
2	0.2	1	*	*	*	*	*	*	*	Allow MORE mismatches in priming region and add --minlength parameters
3	0.1	*	*	*	*	*	*	*	*	Match default but add --discard_untrimmed and --minlength parameters
4	0.4	*	*	*	*	*	*	*	*	Allow MORE mismatches in priming region
5	*	*	*	*	*	*	*	*	32	Allow MORE indels during denoising
6	*	*	*	*	*	*	1e-20	*	*	LESS conservative denoising
7	*	-	*	*	*	*	*	*	*	REMOVE cutadapt minlength requirement
8	*	1	*	*	*	*	1e-40	2	*	REMOVE error correction during denoising, return to other default parameters for clustering, remove 0 length sequences
9	*	*	*	*	*	*	1e-20	*	*	LESS conservative denoising

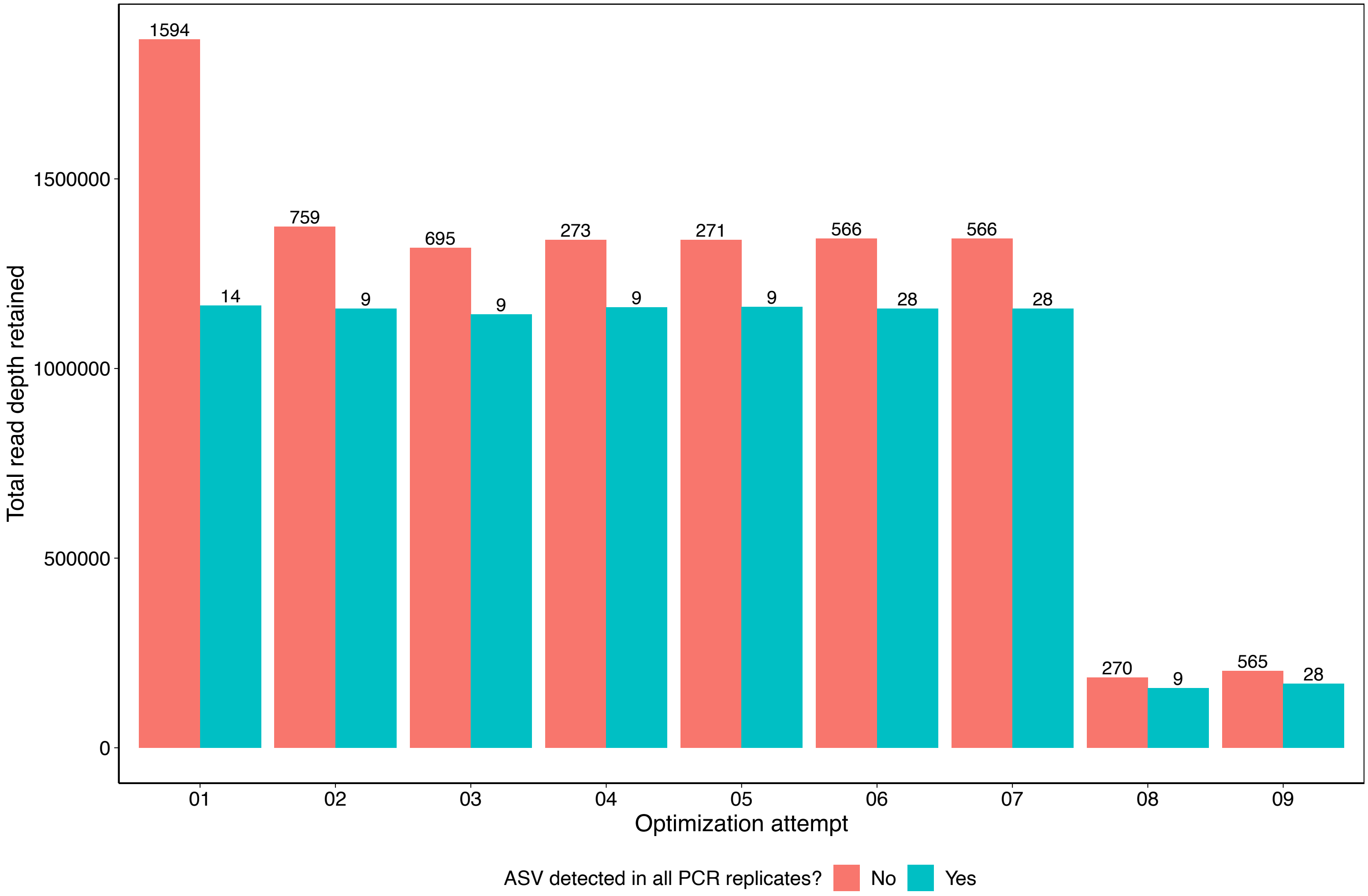

Figure F2 Total read depth retained across DADA2 pipeline optimization attempts. Parameter changes corresponding to the optimization attempt are detailed in Table F2. The value above each bar represents how many ASVs were contained within the retained read depth.

To further confirm the suitability of the optimized pipeline and downstream filtering methods for *Sma*I-corII variant detection, a final comparison was made to DADA2 which included the addition of a 105 bp maximum sequence length enforcement to maintain consistency with the custom pipeline. Following dereplication and applying the same manual depth cut-offs as were applied to the custom pipeline output, DADA2 retained 89 ASVs totalling 301,223 reads. In assessing the ASVs, many were present in high depths in 1/3 PCR replicates of an individual fish and as singletons in the other 2/3 PCR replicates. Of the 89 ASVs, only two were consistently detected at depths greater than 1% in 3/3 PCR replicates and the depth of these two ASVs totalled 144,572 reads. Therefore, the custom pipeline was determined to be better suited for *Sma*I-corII variant determination and the five ASVs retained following analysis with the custom pipeline and optimized downstream depth-based filtering were considered to be putatively true.

Appendix G – Marker comparison on tank dilution series

The sensitivity comparison between *Sma*I-corII SINE and a mitochondrial marker (Chapter 2) was originally planned for a mitochondrial marker targeting one of the common Atlantic Whitefish mitochondrial haplotypes, Chu1, and the universal primer MiFish-U which targets the 12S ribosomal RNA gene (Miya et al. 2015). Due to the inclusion of the 12S marker, DNA sequencing methods were used for this analysis rather than the qPCR methods employed in Chapter 2. Returned read depths for the SINE were lower than expected given the input sequencing parameters and this discovery initiated the optimization of bioinformatic processes described in Chapter 3. The three libraries were prepared with different sequencing depths and therefore direct comparison of read depths across markers was not possible. Instead, sensitivity was determined by where detection was last observed in the dilution series. The SINE marker (run with the optimized analysis determined in Chapter 3) had detection in multiple technical replicates at the lowest dilution of 10^{-5} while the targeted mtDNA and universal mtDNA markers last had detections in the 10^{-4} and 10^{-3} dilutions, respectively (Figure G1).

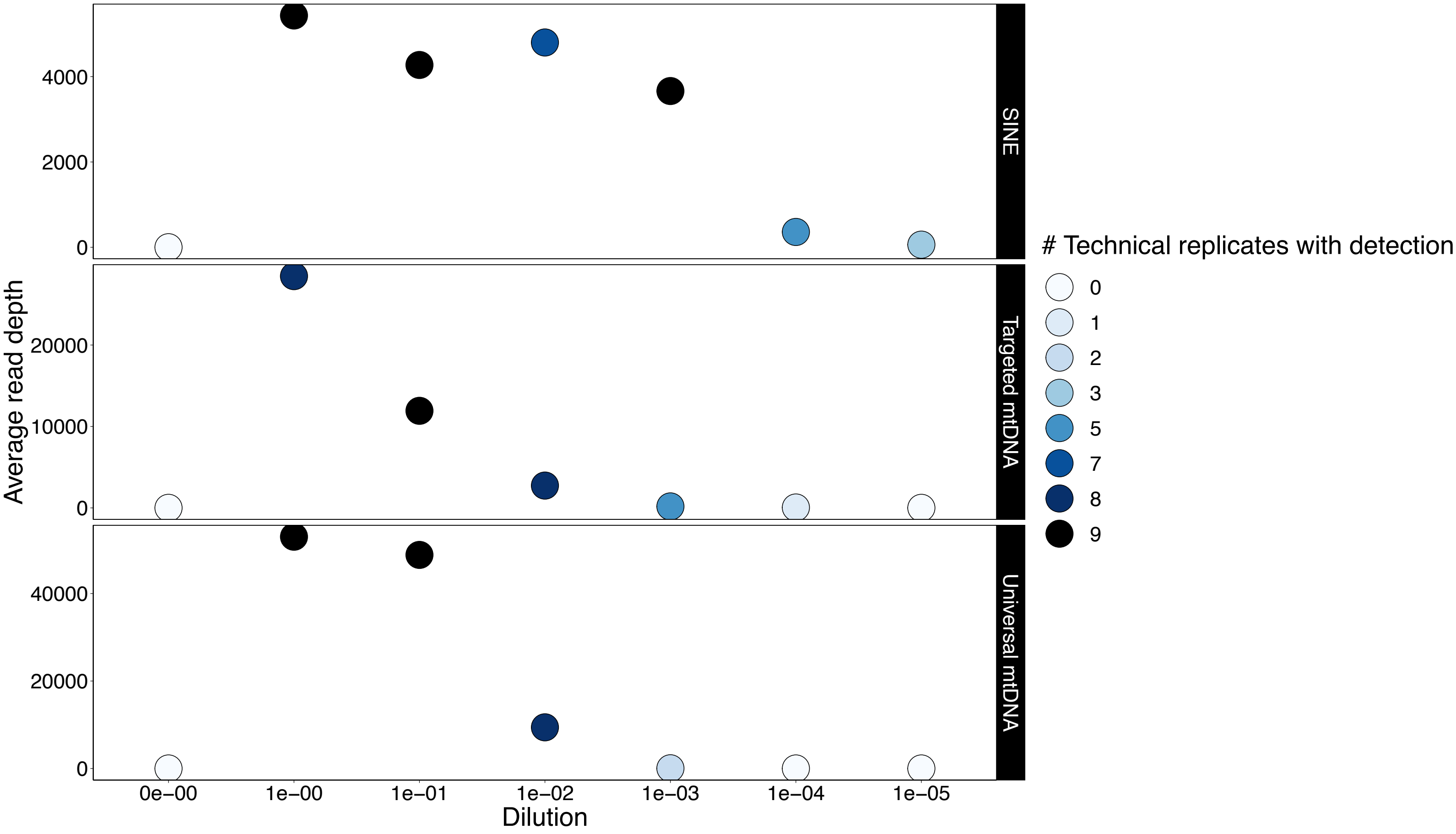

Figure G1 Dilution series marker comparison of the *Sma*I-corII SINE (optimized pipeline) and mitochondrial markers targeting an Atlantic Whitefish mitochondrial haplotype (targeted mtDNA) and 12S rRNA region (universal mtDNA).