Evaluating non-detection risk associated with high-throughput metabarcoding methods for early detection of aquatic invasive species

A Thesis
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Chelsea L. Hatzenbuhler

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Dr. John R. Kelly, Advisor

June 2015

Acknowledgements

Sincere thanks to the UM-EPA Cooperative Training Partnership for giving me the opportunity to complete my graduate study at a world-class government laboratory, and for providing funding to present my results at the 2014 Upper Mid-West Invasive Species Conference. I could not have completed this work without the knowledge and guidance of my academic advisor, John R. Kelly, who helped focus my research into a cohesive and meaningful thesis. My gratitude goes to the other members of my thesis committee, Donn Branstrator and Randall Hicks for your support of my proposal, thesis and your generosity with your time. I also owe much thanks to the outstanding scientists at the Duluth, MN EPA laboratory for their assistance with larval fish sample collection, taxonomy and QA/QC, and to the group at the EPA lab in Cincinnati, who contributed a great amount of time conducting and providing training for the molecular and bioinformatic methods used in this research, as well as for their guidance and feedback regarding the pilot studies and each major experiment we conducted.

Dedication

This project is dedicated to my friends and family, whose constant support has been extremely valuable during my graduate studies.

Abstract

Given the costs associated with traditional taxonomic identification of many aquatic organisms, high-throughput metabarcoding analyses have gained recognition as potentially powerful tools for early detection of aquatic invasive species. A practical early detection strategy, however, demands balancing detection costs with an acceptable level of non-detection risk. Here we evaluated non-detection risk associated with some standard metabarcoding methods by constructing artificial community samples with known species richness and relative biomass abundance composed of fish tissue from multiple “non-target” species and spiked with various proportions “target” tissue from a single species not already present in the sample. Our main findings provided convincing experimental evidence that we can detect the genetic signal produced by target species comprising as low as 0.02% - 1% of total sample biomass and demonstrated the lowest limit of detection observed for each target species varied between experiments.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Overview of aquatic invasive species (AIS) introductions in the Laurentian Great Lakes region

Aquatic invasive species (AIS) have been documented in the Laurentian Great Lakes region beginning in the early 1800s. To date there have been approximately180 intentional and accidental introductions; over 80% were plant (55), invertebrate (55) and fish (36) taxa, many of which successfully colonized and spread throughout the Great Lakes basin (USGS 2012). Anthropogenic vectors associated with live organism trade (Mills *et al.* 1993; Hall & Mills 2000) and the international shipping industry are primary mechanisms for aquatic invasive species (AIS) transport into the Great Lakes region (Mills *et al.* 1993; Ricciardi 2001; Ricciardi 2006). Almost 50% of all introductions have occurred in the past 50 years (USGS 2012) of which 60% were facilitated by ballast discharge from transoceanic vessels (Ricciardi 2006; Pothoven *et al.* 2007). In addition to commercial transport, recreational boating and fishing gear have facilitated the spread of established AIS populations within the region (Mills *et al.* 1993; Ricciardi 2006; Rothlisberger *et al.* 2010).

Successful colonizers (i.e., established AIS populations) endanger the economic and ecological constitution of invaded systems. Ecological changes transpire through a variety of processes such as predation (Krueger & May 1991), parasitism (Schneider *et al.* 1996), interspecific competition (Boileau 1985; Krueger & May 1991) and habitat disturbance (Hecky *et al.* 2004; Zhu *et* *al.* 2006). The ecological impacts caused by unchecked AIS populations often have negative economic consequences; historically, AIS introductions have reduced native commercial, sport and forage fish populations (Smith 1970; Crowder 1980; Krueger & May 1991; Schneider *et al.* 1996). Additional costs arise from remediation efforts associated with water biofouling which impairs industrial and recreational water uses (Mills 1994; MacIsaac 1996). However, the greatest costs relate to management efforts aimed at AIS prevention, control and eradication (Lovell & Stone 2005; Pimentel, Zuniga & Morrison 2005; Lodge *et al.* 2006).

AIS management

AIS management efforts have largely focused on preventing new introductions, controlling AIS spread and eradicating established populations. To reduce AIS introduction (i.e., propagule pressure), 1993 federal regulation mandated transoceanic ships to complete open-ocean ballast water exchange prior to entering the St. Lawrence Seaway which is the water route to the Laurentian Great Lakes (CFR 1993). In addition to U.S. federal regulations, state governments have taken steps to prevent the spread of AIS within the Great Lakes and to inland water bodies (MNDNR 2014). Despite preventative actions, 22 new species have been introduced into the region after ballast water regulations were enacted (USGS 2012). The continued influx and spread of AIS in the Great Lakes region has encouraged extensive development in risk screening and early detection methods which include using new computational software, analytical approaches, as well as implementing the use of new molecular diagnostic tools to increase detection efficiency (Lodge *et al.* 2006; Vander Zanden *et al.* 2010). In particular, consistent monitoring for new invasions can strengthen management success by targeting invasion prone locations and conducting surveysdesigned to detect new AIS during the early stages of the invasion process when specimens are present at low abundance (rare) and the population is localized (Hulme 2006; Lodge *et al.* 2006; Vander Zanden *et al.* 2010). Early detection strategies require substantial time and effort to minimize the probability of not finding a new invader when individuals are present (i.e., a false non-detection event), as the risk of non-detection is inversely proportional to the thoroughness of the search (Hoffman *et al.* 2011). Early detection monitoring coupled with a rapid response to positive AIS detection can increase the success rate of control and eradication efforts as well as reduce the associated costs (Lodge *et al.* 2006).

Overview of AIS early detection strategy

The primary components of detection are: 1) specimen collection, and 2) specimen identification; thus a practical early detection strategy requires achieving balance between efficient sampling (Lodge *et al.* 2006; Trebitz *et al.* 2009; Hoffman *et al.* 2011) and taxonomy methods (Lodge *et al.* 2006) that minimize early detection costs with the risk associated with non-detection for a given detection level. Sampling effort, however, is indirectly related to population size; therefore successful collection of 95 – 100% (i.e., rare – very rare) of species richness estimates for a sampled area entails considerable effort to achieve a high probability of detection and requires a very large quantity of samples/specimens (Trebitz *et al.* 2009; Hoffman *et al.* 2011). Detection efficiency can be improved by modifying collection methods to sample a variety of habitats with multiple sampling gears (Trebitz *et al.* 2010). Although efficient collection methods facilitate early detection, accurate taxonomic identification is critical to minimize non-detection risk and implement AIS management action. There is ongoing research to examine if efficiencies may be gained for detecting invasive fish species through sampling designs aimed at larval specimens instead of adults, primarily because the presence of larvae suggests a reproducing population, which is necessary for successful AIS colonization. However, because collection of larvae results in a very large number of specimens, taxonomy costs increase and the level of non-detection risk associated with morphological taxonomy is uncertain.

Traditional taxonomy exploits morphological diversity to discriminate species. Accurate, high-resolution taxonomy requires substantial time and effort by expertly trained taxonomists, but conflicting expert identifications reduce taxonomic certainty and specimen characteristics can impede resolution and accuracy. Fish undergo four larval stages including yolk-sac, pre-flexion, flexion and post-flexion stages; each life stage requires a different taxonomic key for species level identifications and for some taxa the keys are incomplete or inaccurate (Simon & Vondruska 1991; French III & Edsall 1992) rendering accurate, high resolution taxonomy impossible. Taxonomic uncertainty can also arise when specimens lack identifying features due to damage sustained during field collection. Furthermore, cryptically diverse and unfamiliar or rare species (i.e., newly introduced species) (Hebert *et al.* 2004a; Spies *et al.* 2006; Saunders 2009; Matarese *et al.* 2011) may evade detection due to insufficient taxonomic resolution or specimen misidentification.

Taxonomy methods employed in a practical, successful AIS early detection program must balance associated costs (e.g., time and effort) with minimal risk of non-detection. However, non-detection risk associated with morphological taxonomy is unknown and may change between studies, as positive species level detection is dependent on many variables (e.g., expertise, key availability/accuracy, specimen condition/life stage) (Stribling *et al.* 2008; Haase *et al.* 2010). Given that non-detection risk is uncertain, an alternative to traditional taxonomy is needed for early detection strategies. Recent technological advances in molecular biology (e.g., DNA barcoding, high-throughput sequencing) techniques provide fast, accurate and cost-effective methods for species identification (Ji *et al.* 2013; Ko *et al.* 2013; Stein *et al.* 2014) that hold promise for an AIS early detection strategy; moreover, the associated limits of detection are quantifiable and can be highly sensitive in the sense of generating genetic data for low abundance taxa (Hajibabaei *et al.* 2011; Pochon *et al.* 2013; Zhan *et al.* 2013).

AIS early detection and DNA based methods for species identification

Generally, the molecular taxonomic methods used in AIS early detection strategies are described as a targeted or community approach to detection. Targeted techniques generate genetic presence/absence data for a single predetermined target species, but full species composition is not determined (Bott *et al.* 2010; Jerde *et al.* 2011; Goldberg *et al.* 2013; Jerde *et al.* 2013; Mahon *et al.* 2013). Conversely, the community approach to detection (i.e., metabarcoding) does not require prior knowledge of the target species. This approach uses high-throughput sequencing technology that enables simultaneous sequencing of many samples (i.e., multiplexing) to generate sequences from a genetic marker to determine sample richness.

Using metabarcoding methods for AIS early detection may facilitate species level identifications, but selecting an appropriate marker is crucial toattain this high-resolution taxonomy and maximize detection efficiency. The genetic marker, or DNA barcode promotes species discrimination because interspecific genetic variation exceeds intraspecific variation. In addition, the ends, or flanking regions of barcodes consist of nucleotide bases that are highly conserved between taxa. This characteristic allows for a universal primer design that enables routine recovery & PCR amplification of barcodes in multi-species assemblages (Folmer *et al.* 1994; Ward *et al.* 2005; Ivanova *et al.* 2007).

The mitochondrial genome (mtDNA) is particularily well suited for DNA barcoding as the genome size, shape and gene content are highly conserved for most animal phyla. The mode of inheritance is haploid so genetic recombination is limited therefore intraspecific variation is low (Meyer 1993; Hebert *et al.* 2003), yet interspecific sequence diversity is high (Hebert, Ratnasingham & deWaard 2003) due to an increased molecular rate of evolution relative to nuclear DNA (Saccone *et al.* 1999). There can also be multiple copies of mtDNA per cell (Meyer 1993), so genes are readily amplified which may allow rare species to be more easily detected. In addition, most mtDNA protein coding genes lack insertions and deletions (Meyer 1993) that can complicate the sequence alignments necessary for assigning taxonomy to unidentified barcodes (Ratnasingham & Hebert 2007).

In recent years, the 650 bp section off the 5' end of the mtDNA protein coding gene, cytochrome *c* oxidase (CO1) was designated as a standard DNA barcode for many animal groups (Hebert, Ratnasingham & deWaard 2003; Hebert *et al.* 2004a; Hebert *et al.* 2004b) including marine and freshwater fish taxa (Ward *et al.* 2005; Ko *et al.* 2013). In addition, the establishment of standard DNA barcodes like CO1 has prompted a global effort to develop a publicly available, comprehensive reference sequence database which is crucial for assigning taxononomy to unknown barcodes (Ratnasingham & Hebert 2007). Reference sequences are linked to adult voucher specimen taxonomy supported by expert morphological identifications (Meyer & Paulay 2005). Unidentified barcodes are assigned taxonomy when aligned to a similar reference sequence (Benson *et al.* 2005; Ratnasingham & Hebert 2007). After taxonomy is assigned, richness can be determined for the original samples; but final richness estimates and ultimately positive species level detection depends on the taxonomic resolution of the selected genetic marker (Hebert *et al.* 2003; Hajibabaei *et al.* 2011; Pochon *et al.* 2013; Zhan *et al.* 2013). So, a high-resolution marker, like CO1, and a comprehensive reference database is necessary to minimize non-detection risk associated with metabarcoding approaches for AIS early detection.

Metabarcoding sensitivity and potential sources of non-detection risk

A few recent studies have demonstrated high-throughput metabarcoding methods for species richness determination to be very accurate as well as highly sensitive in constructed aquatic invertebrate community samples with known biodiversity. For samples constructed from benthic macro-invertebrate tissues, CO1 barcodes were detected for all taxa present at > 1% relative biomass abundance (Hajibabaei, Shokralla et al. 2011). An even lower detection level was achieved at > 0.64% relative biomass abundance for samples constructed using DNA extracted from marine invertebrate tissues to generate sequences for genes from nuclear small subunit ribosomal DNA (Pochon *et al.* 2013). If biomass abundance corresponds to the genetic signal, (the total sequences, or relative sequence abundance produced per taxa for each sample), then signals produced by low abundance, or rare taxa, should be represented by fewer sequences relative to more common species and extremely rare taxa would be represented by only one or two sequences (Zhan *et al.* 2013). But in samples with unknown biodiversity, genetic signals represented by very few sequences (weak signals), are generally associated with sequencing errors that may lead to false positive detection. To reduce the likelihood of false positive detection events (Kunin *et al.* 2010), standard data processing methods are used to denoise sequence data, by removing low quality and potentially erroneous, or biologically irrelevant, sequences (Caporaso *et al.* 2010). Since biological relevance is unknown, weak signals are also filtered out and excluded from final sequence biodiversity measurements. Denoising increases the risk of non-detection because filtered sequences may be biologically relevant; however, the utility of retaining or removing sequences associated with weak signals for studies focusing on rare species detection is unclear (Zhan *et al.* 2014).

In addition to sequence data processing, upstream processes within the complex sequencing workflow, as well as factors pertaining to the experimental design, sample collection and processing also have potential to affect detection (Fig. 1). For instance, sample collection and processing methods can influence the quality of DNA used for PCR (polymerase chain reaction) amplification of DNA barcodes. In addition, differential barcode amplification (PCR bias) resulting from random amplification (PCR drift), interspecific variation in primer binding affinity or gene copy number can skew sequence biodiversity estimates from corresponding biomass abundance (Wagner *et al.* 1994; Polz & Cavanaugh

1998) and extreme biases may lead to false non-detection events (i.e., detection errors) for under-represented or rare taxa; however, barcode selection and PCR design can reduce bias intensity. In addition, sample composition, in conjunction with workflow processes, may also increase the risk of non-detection. The inherent variation observed in aquatic field samples results in varying levels of complexity. Life stage or biodiversity may vary for sampled taxa within and between samples, but it is currently unknown if and how sample complexity influences detection success. Furthermore, aquatic samples can contain sample residue comprised of detritus, sediments and other non-target organisms (e.g., invertebrates present in larval fish samples). Sample processing includes separating sampled taxa from residue prior to traditional taxonomy and the time and effort required for this step is dependent on the amount and type of residue. Exploring the effects of sample residue (referred to hereafter as *detritus*) on detection can provide insight into the extent of sample processing necessary for metabarcoding analysis. Sensitivity assessements (Hajibabaei *et al.* 2011; Pochon *et al.* 2013; Zhan *et al.* 2013) have shown the lowest limits of detection associated with metabarcoding to be highly sensitive for some aquatic invertebrates. For fish communities, a relevant biological example for AIS early detection strategies, the lowest limits of detection have not yet been reported and current understanding relative to the risk of non-detection associated with workflow processes is inadequate; consequently the utility of metabarcoding methods for AIS early detection remains in question.

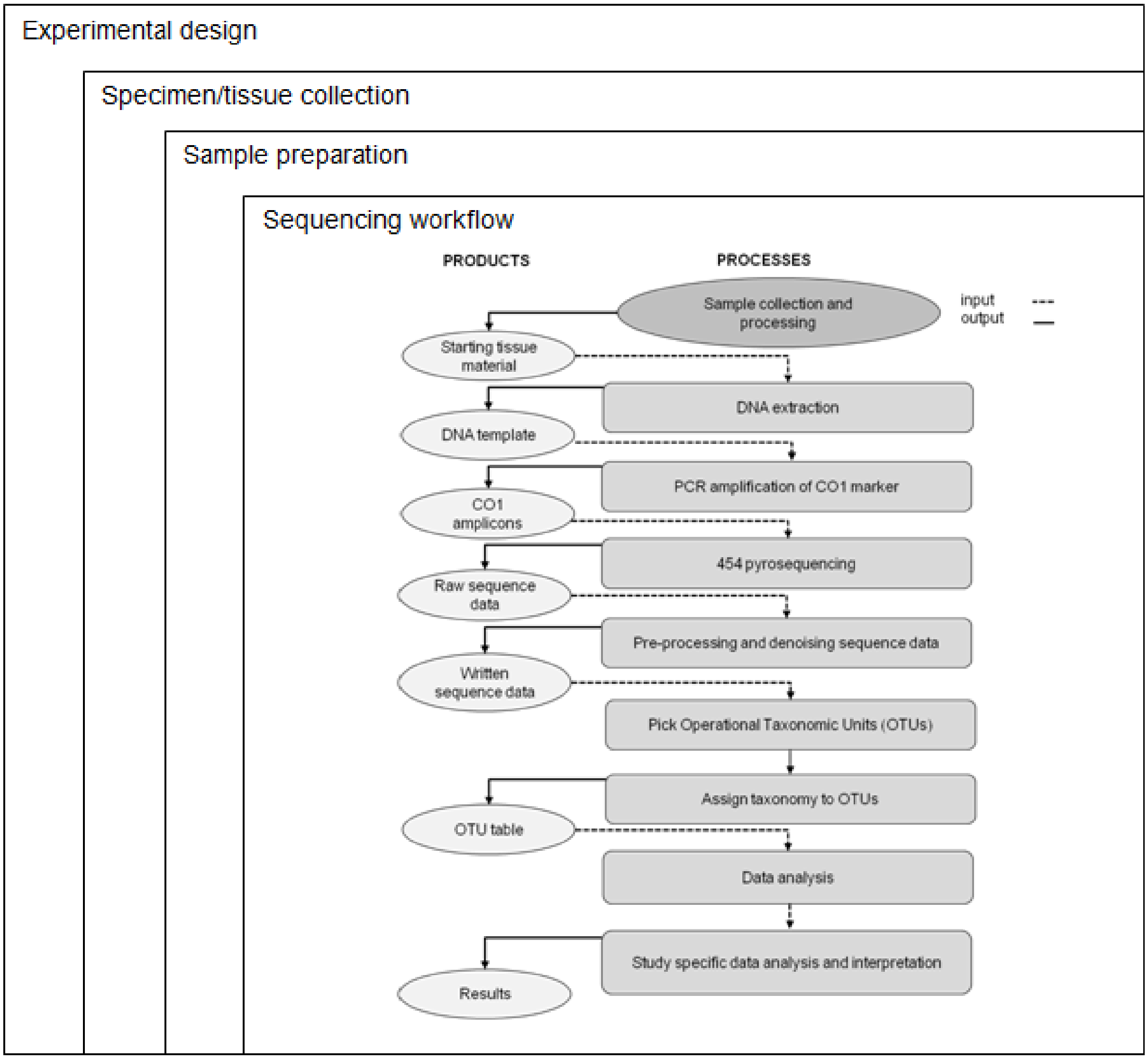


Figure 1 Workflow overview. There are many steps involved within and beyond the sequencing workflow and methods used in each stage may influence positive detection of low abundance taxa in multi-species assemblages (dashed lines represent inputs and solid lines represent outputs/products).

Research objectives

To shed light on detection limits associated with commonly used metabarcoding methods for species identification, we carried out several experiments designed to investigate detection sensitivity and the effects of sample complexity on species level detection in constructed larval fish community samples. We chose larval fish as a relevant life stage because traditional taxonomy for larval fishes can be very challenging. Furthermore, in general the risk of non-detection for traditional taxonomic methods is uncertain and is not easily reduced. Therefore, metabarcoding as an alternative to traditional taxonomy may improve the efficiency of an AIS early detection program, if the related non-detection risk low or is reducible. Our pilot study was designed to determine a testable range of biomass based detection probabilities and define workflow processes that influence detection success. Pilot results directed modifications made to the design, sample construction methods and/or workflow methods used for subsequent experiments that aimed to assess non-detection risk under circumstances designed to maximize detection success. We suspected the sensitivity afforded by high-throughput sequencing technology for detection of low abundance fish species in constructed community samples to reflect or exceed sensitivity described for some aquatic invertebrate taxa (Hajibabaei *et al.* 2011; Pochon *et al.* 2013; Zhan *et al.* 2013). Based on our assessment of non-detection risk associated with sequencing workflow processes, we expected to gain insight into methods in need of modifications to optimize detection of fish species.

METHODS

Experimental design

To evaluate the limits of detection, several experiments were conducted using constructed community samples comprised of fish tissue. Constructed sample matrices contained a mixture of tissue from taxa classified as a “non-target” or “target” species and for each experiment, a single target species was selected to represent the rare taxa Non-detection risk was assessed in a range of sample matrices. Sample composition was manipulated to increase matrix complexity related to sources of inherent variation, such as species richness and inclusion of detritus. Standard methods were used to analyze the first experimental sample set and clarify sources of detection error (i.e., workflow processes causing false non-detection events); design and methods for subsequent sample sets were modified to minimize detection error. Each experiment was designed to limit the potential for detection errors related to sample matrix composition and evaluate our ability to detect low abundance taxa. To limit detection error related to genetic distance (divergence) we selected taxa from distinct genera or families (distantly related taxa). We attempted to reduce the effects of differential gene copy number on PCR amplification by selecting specimens from similar life stages and homogenizing tissues when multiple specimens were needed from the same species.

A suite of samples was constructed for each experiment; for each set there were three controls including an individual control for each species, a non-target mix (equal biomass proportions of each non-target, target excluded), and an equal proportion control (equal biomass proportions from each non-target and target species). Experimental mixes were comprised of a non-target mix with equal proportions of non-target biomass; non-target matrices were spiked with target tissue to achieve a specific biomass ratio where target biomass varied in decreasing proportion to total sample mass (e.g., biomass ratios equal to 1:5, 1:100 and 1:1000 corresponded to target tissue representing 20%, 1% and 0.1% of total sample biomass, respectively, and ratios reflected the target detection levels tested in each experiment).

The pilot study, experimental sample Set 1 (S1) was constructed with a simple sample matrix, based on the average species richness observed in larval field collected samples; matrices were comprised of unexposed, internal adult fish muscle tissue from five specimens in distinct families (Appendix A, Table A.1). At the expense of replication, S1 was designed to test a wide range of detection levels to determine an appropriate range to test in subsequent experiments. Standard metabarcoding methods were used to define sources of detection error that were investigated and modified as necessary to reduce error in the next experiment. Design for experimental sample Set 2 (S2) mirrored S1 with modifications made to reduce detection error related to sample construction methods and advance our non-detection risk evaluation. S2 was constructed with a simple sample matrix using larval fish tissue homogenates prepared from five fish taxa different from S1 (Appendix A, A.1). Replication was increased in S2 experimental mixes, which tested a more restricted range of detection levels that reflected observations, made from S1 sequence data analysis (Table 1a, b). Experimental sample Set 3 (S3) was constructed to assess non-detection risk associated with increased sample matrix complexity related to species richness and inclussion of detritus using larval and detrital tissue homogenates. For Treatment 1 (S3T1) species richness was manipulated to evaluate target detection in three conditions and samples were constructed with two (S3T1-a), five (S3T1-b) or eleven (S3T1-c) taxa (Appendix A, A.1). Although restricted relative to S2 the range of detection levels tested in each condition (a – c) was constant (Table 2a). For the second Treatment, (S3T2) evaluating the effects of detritus presence on detection, the same target and non-target taxa were used as in S3T1-b (Appendix A, A.1). The ratio of detritus to fish tissue was varied but the ratio of target to non-target biomass remained constant where target biomass represented 1% of total fish mass (Table 2b).

Field sample collection/biomass accumulation

Fish were collected from Lake Superior and its tributaries to provide tissue needs for experimental sample construction. For S1 frozen adult specimens were

Table 1 Summary of experimental design for **a.** Set 1 (S1) and **b.** Set 2 (S2). Relative biomass abundance per taxa as a percent of total sample biomass for controls (not listed in the table are individual controls which were constructed for each taxa), and experimental test mixes assessing our ability to recover sequences for the target species across a range of detection levels. *Replicates were constructed but some were not sequenced because of pre-sequencing complications.

		Relative biomass abundance %				
	Replicates	Target	Non-target			
a. S1		*O. mordax*	*Catostomus spp.*	*L.lota*	*P. nigromaculatus*	*P. flavescens*
Controls						
Equal proportion non-target mix	2	-	25	25	25	25
Equal proportion 1:5	2	20	20	20	20	20
Experimental mix: ratio of target biomass to total sample mass						
1:10	2*	10	22.5	22.5	22.5	22.5
1:100	2*	1	24.75	24.75	24.75	24.75
1:1000	3*	0.1	25	25	25	25
1:5000	3	0.02	25	25	25	25
1:10000	3	0.01	25	25	25	25
1:50000	3*	0.002	25	25	25	25
b. S2		*P. semilunaris*	*A. rupestris*	*E. lucius*	*G. aculeatus*	*N. hudsonius*
Controls						
Equal non-target mix	2	-	25	25	25	25
Equal proportion 1:5	4	20	20	20	20	20
Experimental mix: ratio of target biomass to total sample mass						
1:1000	4	1	25	25	25	25
1:2500	4	0.04	25	25	25	25
1:5000	4	0.02	25	25	25	25

obtained from research conducted at USEPA-MED, Duluth, MN in 2010 – 2011. For S2, S3T1 and S3T2 larval fish were collected from the St. Louis River estuary and Duluth-Superior harbor during June and July 2013. Larvae were sampled following USEPA sampling gear standard operating protocols for each gear type which included 500 um mesh larval tucker trawl, tow sled, beach seine and dip nets. In total 55 samples were collected (23 tucker trawl, 23 dip net, 7 beach seine, 2 tow sled samples). Additional larval specimens were obtained as needed from a 2013 USEPA larval fish survey also conducted on the St. Louis River estuary. Detritus used for S3T2 sample construction was taken from larval fish field samples and included a mixture of filamentous algae, woody debris, decaying organic matter, and aquatic plants; invertebrates, soil particles, sand, manmade materials Environmental DNA, or eDNA, (i.e., sloughed cells, scales, tissue or excretions) from fish could also have been present.

Larval fish processing and taxonomy

DNA degradation and sample contamination was prevented by following DNA sample handling and preservation protocols for all field and laboratory processes. Tissues were preserved in 95% ethanol (EtOH) at the time of collection and stored at or below 4°C (Prendini, Hanner & DeSalle 2002; King & Porter 2004; Nagy 2010; Jackson *et al.* 2012; Stein *et al.* 2013). Laboratory glassware and stainless steel tools were disinfected in a 10% bleach solution for ≥ 10 minutes (Arena 2010) or with the surface contaminant remover DNA

Table 2 Summary of experimental design for Set 3 designed to assess non-detection risk associated with increased sample matrix complexity. Relative biomass abundance per taxa as a percent of total sample biomass for controls (not listed in the table are individual controls which were constructed for each taxa and experimental test mixes assessing our ability to recover sequences for the target species across a range of detection levels in: **a.** Treatment 1 (S3T1) which evaluated complexity related to species richness in samples with two (S3T1-a), five (S3T1-b) and eleven taxa (S3T1-c) and **b.**Treatment 2 (S3T2) which evaluated complexity related to detritus presence with five taxa where target detection level was constant for all test mixes, but the ratio of detritus to fish biomass was manipulated to reflect an increased amount of detritus.

a.S3T1		Target	Non-target taxa									
	Replicates	*P. omiscomaycus*	*Catostomus spp.*	*P. flavescens*	*E. nigrum*	*P. semilunaris*	*M. salmoides*	*A. rupestris*	*P. caprodes*	*O. mordax*	*N. crysoleucas*	*E. Lucius*
S3T1-a												
Controls												
Equal proportion	4	50	50									
Experimental mix: ratio of target biomass to total sample mass												
1:100	4	1	99									
1:300	4	0.33	99.67									
1:600	4	0.167	99.83									
1:800	4	0.125	99.88									
S3T1-b												
Controls												
Equal proportion non-target mix	2	-	25	25	25	25						
Equal proportion	4	20	20	20	20	20						

Table 2 (continued)

Experimental mix: ratio of target biomass to total sample mass												
1:100	4	1	24.75	24.75	24.75	24.75						
1:300	4	0.33	24.92	24.92	24.92	24.92						
1:600	4	0.167	24.96	24.96	24.96	24.96						
1:800	4	0.125	24.97	24.97	24.97	24.97						
1:2000	4	0.05	24.99	24.99	24.99	24.99						
S3T1-c Control Equal proportion non-target mix	2	-	10	10	10	10	10	10	10	10	10	10
Equal proportion	4	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09	9.09
Experimental mix: ratio of target biomass to total sample mass												
1:100	4	1	9.9	9.9	9.9	9.9	9.9	9.9	9.9	9.9	9.9	9.9
1:300	4	0.33	9.97	9.97	9.97	9.97	9.97	9.97	9.97	9.97	9.97	9.97
1:600	4	0.167	9.99	9.99	9.99	9.99	9.99	9.99	9.99	9.99	9.99	9.99
1:800	4	0.125	9.99	9.99	9.99	9.99	9.99	9.99	9.99	9.99	9.99	9.99

Table 2 (continued)

			Target	Non-target					
b.. S3T2	Detritus: fish ratio	Replicates	*P. omiscomaycus*	*Catostomus spp.*	*P. flavescens*	*E. nigrum*	*P. semilunaris*	Percentage of total sample mass	
Control								Detritus	Fish
Equal proportion	0:1	4*	20	20	20	20	20	-	100
Experimental mix: ratio of target biomass to total sample mass									
-	1:0	4*	-	-	-	-	-	100	-
1:100	0:1	4*	1	24.75	24.75	24.75	24.75	-	100
1:100	1:1	4*	1	24.75	24.75	24.75	24.75	50	50
1:100	3:2	4*	1	24.75	24.75	24.75	24.75	60	40
1:100	1:2	4	1	24.75	24.75	24.75	24.75	33.3	66.6
1:100	1:10	4*	1	24.75	24.75	24.75	24.75	9	91
1:100	1:20	4	1	24.75	24.75	24.75	24.75	5	95

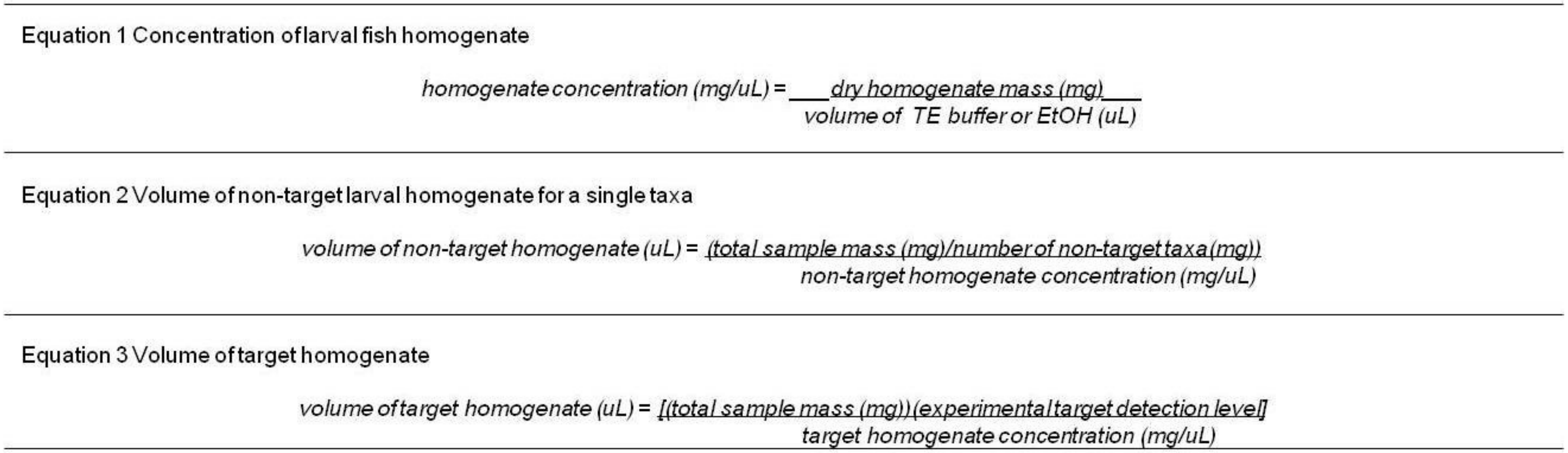
away™, rinsed with sterile water followed by a final EtOH rinse prior to air drying. Aluminum weigh pans were heat sterilized ≥ 24 hrs at 180°C. Larval fish specimens were removed from sample residue and re-preserved in 95% EtOH; larvae were identified (Auer 1982) and quantified, then pooled based on taxonomy and larval life stage. A second larval fish taxonomist verified the taxonomy of larvae selected for sample construction.

Tissue preparation and sample construction

Internal muscle tissue used to construct S1 samples was removed from frozen adult specimens to prevent interspecific DNA contamination. DNA quality was determined prior to sample construction; genomic DNA was extracted from muscle tissue using DNeasy ® Blood and Tissue Kit (Qiagen, Hilden, Germany) and DNA extraction success (i.e., good DNA quality & high quantity) was verified by electrophoresis with 1% agarose gel. Muscle tissue was extracted from well-preserved specimens and re-preserved in 95% EtOH. Excess EtOH was blotted from muscle tissue to obtain a more accurate tissue mass measurement for controls and test sample replicates. Complete, constructed samples were re-preserved in 95% EtOH and stored in sterile glass scintillation vials at ≤ 4°C until submitted into the sequencing workflow (Fig. 1).

Larval fish tissue homogenates were made from each species for S2 and S3T1, S3T2 samples; in addition, a detritus homogenate was made for S3T2. Tissue mass was placed into a -20°C chilled mortar, immersed in liquid nitrogen, and homogenized according to the USEPA standard operating procedure for larval fish tissue homogenization for DNA analysis (Martinson & Struewing 2010; Burden 2012). The cryogenic homogenate was desiccated (Nagy 2010) and desiccated homogenate was weighed. For S2 and S3T1, 4°C Tris EDTA buffer, pH 8 was added to desiccated homogenate; 95% EtOH was used for S3T2, and for each experiment tissue was re-homogenized using PT-10735 Polytron® Stand Homogenizer (Kinematica AG, Lucerne, Switzerland) (Martinson & Struewing 2010). Equations 1-3 listed in Table 3 were used to calculate homogenate concentration, mg/uL (Table 3, *equation* 1) and homogenate volume (Table 3, *equations 2, 3*). Homogenate aliquots (± 0.00076 mg/uL) contained enough detrital or larval tissue from each taxon to construct samples with biomass ratios specific to detection levels tested in each experiment (Table 1b; 2a, b). Measurement error (± 0.00076 mg/uL) represents the mean absolute deviation, which is the absolute value of the mean difference between expected biomass and observed biomass calculated from 20 replicates each containing 100 uL of 0.236 mg/uL larval tissue homogenate constructed specifically to calculate pipette measurement error. Homogenates were pipetted into 2 or 7 mL polypropylene collection tubes and stored at ≤ 4°C until submitted into the sequencing workflow.

Table 3 Equations (eq.) used to calculate tissue homogenate concentrations (*eq. 1*) and volumes for non-target (*eq. 2*) and target (*eq. 3*) taxa in Set 2 and Set 3 experimental samples. For *eq. 1,* dry homogenate mass refers to the desiccated, cryogenic homogenate. For *eq. 2* total samplemass and number of non-target taxa and for *eq. 3* total sample mass and experimental target detection level are part of the experimental design; target detection level is probability of detection for the target at the corresponding ratio of target mass to total sample mass.



Equation 1 Concentration oflarval fish homogenate

$$\textit{homogenate concentration (mg/uL)} = \frac{\textit{dry homogenate mass (mg)}}{\textit{volume of TE buffer or EtOH (uL)}}$$

Equation 2 Volume of non-target larval homogenate for a single taxa

$$\textit{volume of non-target homogenate (uL)} = \frac{\textit{(total sample mass (mg)/number of non-target taxa(mg))}}{\textit{non-target homogenate concentration (mg/uL)}}$$

Equation 3 Volume of target homogenate

$$\textit{volume of target homogenate (uL)} = \frac{\textit{[(total sample mass (mg))(experimental target detection level]}}{\textit{target homogenate concentration (mg/uL)}}$$

Sequencing workflow: DNA extraction

Total genomic DNA was extracted according to the manufacturer's instructions from vacuum desiccated tissue mass using the PowerMax ® Soil DNA Isolation Kit (MO BIO Laboratories, Inc, Carlsbad, CA) for S1 samples with total mass ≥1000 mg and the DNeasy ® Blood and Tissue Kit (Qiagen, Hilden, Germany). For S2 and S3T1, T2 when total mass ≤ 50 mg. The original extraction protocol for S1 was modified to include an additional 500 uL of proteinase K and an overnight digestion at 56°C to ensure complete deproteinization. DNA extraction success was verified by electrophoresis with 1% agarose gels. Template DNA was quantified using the PicoGreen protocol with Quant-iT ™ PicoGreen ® dsDNA assay (Life Technologies, Carlsbad, CA) and Synergy ™ HT Multi-Mode Microplate Reader (Bio-Tek, Winooski, VT). DNA of acceptable quality and quantity (i.e., large quantity of DNA strands ≥ 400 bp) was normalized using sterile water to 10 ng template DNA/uL.

Polymerase chain reaction

The CO1 markers in the single and multi-template DNA (i.e., DNA template comprised of one or multiple species, respectively) were PCR amplified which produced CO1 amplicons by successive heating and cooling, or thermocycling, of PCR reagents (Appendix A, Table A.2a), CO1 specific primers (Appendix A, Table A.3)) and template DNA. The thermocycler program used in our study aimed to reduce amplification bias associated with PCR drift by amplifying template DNA for each sample in replicates of five and pooling the replicates before PCR purification (Polz & Cavanaugh 1998). PCR reactions took place in a Bio-Rad™ thermocycler and initiated at 94°C for 150 sec., followed by 34 cycles at 94°C for 30 sec., 46°C for 60 sec., and 72°C for 60 sec., then a final extension at 72°C for 10 min. PCR success was verified by electrophoresis on 1% agarose gel. PCR product was purified using Qiagen's PCR Purification Kit to remove DNA fragments smaller than approximately 300 bp. Purified amplicon DNA was quantified using the PicoGreen protocol with Quant-iT ™ PicoGreen ® dsDNA assay (Life Technologies, Carlsbad, CA) and Synergy ™ HT Multi-Mode Microplate Reader (Bio-Tek, Winooski, VT), then normalized to 10 ng amplicon DNA/uL.

Sequencing workflow: 454 pyrosequencing

454 pyrosequencing processes were carried out by our collaborators at USEPA, Cincinnati, OH on the 454 GS-FLX+ ™ instrument per manufacturer's instructions. The 454 instrument was suited for sequencing the CO1 barcode as it generates sequences (reads) up to 1000 bp in length with 99.9% accuracy and can produce an estimated one million reads per run depending on sample plating strategy (454 Life Sciences) (Appendix B.1). The pyrosequencing process began with basic multiplexed amplicon library preparation where 454 fusion primers composed of library adapters, sequencing key and unique molecular identifier tag

(MID tag) that links barcode reads to original sample IDs, were ligated to the CO1 amplicons via a secondary round of PCR. PCR reagents (Appendix A, Table A.2b) were thermocycled under the same conditions described for the initial PCR. The resulting adapter carrying DNA was hybridized with complementary adapters carried by capture beads and amplified via emulsion-based clonal amplification (emPCR). During emPCR, hybridized beads were immersed in solution containing oil and PCR reagents and the solution was emulsified to isolate each bead. PCR reactions resulted in exponential amplification of a single sequence on each bead and the resulting clonal amplicon library carrying beads were centrifuged with sequencing reaction enzymes onto a 70x75 PicoTiter ™ plate (PTP) according to the PTP layout specified for each sample Set (Appendix B.1) and pyrosequenced. During a sequencing run, deoxyribonucleotides (dNTPs) flow over the loaded PTP and complemented nucleotide bases produce a chemi-luminescent signal that is recorded. Signals were visualized in standard pyrosequencing output files, or flowgram files, used for DNA sequence determination during the base calling process. Called sequences were written to the standard bioinformatics format, FASTA, and mapping files were generated to link MID tags to the original sample IDs (454 Life Sciences).

Sequencing workflow: Pre-processing and denoising sequence data

Bioinformatics analysis was carried out by collaborators at the USEPA, Cincinnati, OH. Qiime software (Caporaso *et al.* 2010) was used to process all 454 sequence data. Sample IDs were assigned to multiplexed sequences using the MID tag, mapping file and extraneous sequence data (e.g., primers, library adapters) were trimmed from CO1 sequences. The raw sequence data was pre-processed to filter for quality; sequences below quality filtering thresholds were removed from downstream processes and the remaining, acceptable sequences were written to a new FASTA file (Appendix A, Table A.4). Written sequences were denoised using the Acacia denoising algorithm (Bragg *et al.* 2012) to correct pyrosequencing errors resulting from inaccurate determination of homopolymer length, (i.e., subsequence of identical bases) during the base calling process. Blast_fragments program was used to identify PCR artifacts (chimeras) that result when DNA from two or more species combines to form a single sequence during PCR (Appendix A, Table A.5). Standard metabarcoding analyses remove chimeras to reduce the probability of identifying a false novel organism in sample with unknown richness. For our study, samples were constructed and richness was known, so chimeras were identified and isolated for further investigation into the taxonomic composition of each chimeric sequence.

Sequencing workflow: Assigning Operational Taxonomic Units (OTUs) and Taxonomy

An operational taxonomic unit (OTU) is defined by the taxonomic group being studied (e.g., species, genera, family). UCLUST software (Edgar 2005) was used to cluster sequences at ≥ 97% sequence similarity into OTUs and a single sequence, or representative OTU, was selected to represent the cluster in downstream analyses and linked to multiplexed sequence data (Appendix A, Table A.6). Taxonomy was assigned to representative OTU sequences using the Basic Local Alignment Search Tool (BLAST) algorithm (Altschul *et al.* 1990). OTUs were aligned with sequences from a reference library database comprised of publicly available CO1 sequences downloaded from the Barcode of Life Database (BOLD) (Ratnasingham & Hebert 2007) and CO1 voucher sequences from fish specimens collected in the Great Lakes basin by the USEPA Duluth, MN, U.S. Fish and Wildlife Service, Minnesota and Wisconsin Departments of Natural Resources. Taxonomy assignments were based on percent match criteria of > 90% similarity to a reference sequence. All bioinformatics files were sent to USEPA, Duluth, MN where several study specific data analyses were conducted prior to data interpretation.

Sequencing workflow: Study specific data analysis and interpretation

For each experiment species richness was compared between the constructed samples and corresponding sequence data. Detection success was assessed for non-target and target taxa in controls and experimental mixes. Positive detection was indicated by the presence of sequences associated with biologically relevant expected values or e-value (the probability of taxonomic alignment occurring by chance, reflecting the biological relevance of taxonomic assignments) . The lowest limit of detection (LLD) for target was associated with the sample containing the smallest proportion of target biomass where target signal was positively detected. For each non-target and target species detection success rates were calculated for samples with replication where rates were the percent of replicates in which the genetic signal was observed. Finally, for each sample, sequences were converted to relative sequence abundance per taxa and compared to the corresponding relative biomass abundance using a Pearson's chi squared test of independence to determine the significance of observed differences ($\alpha = 0.05$, $\alpha = 0.001$) (Table 4); a Yates correction for continuity was applied to S3T1-c as expected values for the test were < 10 mg.

Several study specific analyses were conducted to investigate the effect of weak signal removal on detection and determine the legitimacy of perceived non-detection events. OTUs assigned low-resolution taxonomy (i.e., taxonomy other than species level) were realigned to sequences in the larger online version of BOLD (Ratnasingham & Hebert 2007) and GenBank ® (Benson *et al.* 2005) databases to attain higher resolution identifications if possible. Isolated chimeric sequences were also analyzed to validate non-detection events. Collectively, these steps functioned to recover sequences, or amplify signals for species that were initially undetected due to flaws in the reference sequence database causing false non-detection events.

After sequence realignments, the associated e-values and signal strength (number of clustered sequences) of OTUs associated with low-resolution or unexpected fish taxonomy (i.e., fish species not used to construct samples) served to set threshold values for weak signal removal, which were set for each sample. All sequences, including sequences associated with expected fish taxonomies (i.e., fish species used to construct samples), falling below threshold values were removed from the final data set. This filtering process, or weak signal removal is common in sequence data analyses and methods used to set threshold values are relative to the study design. For our study, thresholds were set based on our preexisting knowledge of sample composition. Therefore, in addition to sequencing errors, sequences were filtered to reduce the probability that false positive detection events be attributed to DNA contamination occurring during sample construction.

To gain insight into the effect weak signal removal on detection rates and the LLD, replicates in each sample set were independently analyzed before and after filtering for weak signals. Results for the LLD and detection rates were reported for both the unfiltered and filtered data and in cases where sequences were recovered after OTU-realignment, the unfiltered and filtered datasets were analyzed before and after recovering sequences. In total there were four possible data sets to analyze for each experiment; unfiltered and filtered data analyzed before and after sequence recovery, unfiltered/before (UB); unfiltered/after (UA); filtered/before (FB); filtered/after (FA). When realignments resulted in zero

Table 4 Pearson's Chi squared test results for **a.** non-target and target taxa (shaded) in the equal proportion control replicates for experimental sample sets 1 (S1), 2 (S2) and 3, treatment 1 (S3T1) where sample matrices were constructed with two (S3T1-a), five (S1, S2, S3T1-b) or eleven (S3T1-c) taxa. **b.** non-target taxa in S1, S2, S3T1-a, b, c replicates with equivalent non-target relative biomass abundance, where target biomass represented ≤ 10% of total sample mass.

		a. Equal proportion controls			**b.** Experimental mixes		
		df	X^2	Significance α = 0.05* , 0.001**	df	X^2	Significance α = 0.05* , 0.001**
S1							
	L. lota	1	307.791	**	12	609.835	**
	P. nigromaculatus	1	36.597	**	12	100.724	**
	P. flavescens	1	9.817	*	12	75.082	**
	Catostomus spp.	1	36.423	**	12	259.779	**
	O. mordax	1	10.463	*	-	-	-
S2							
	N. hudsonius	3	125.024	**	12	406.797	**
	E. lucius	3	6.231	-	12	58.481	**
	G. aculeatus	3	0.213	-	12	19.512	-
	A. rupestris	3	7.029	-	12	73.826	**
	P. semilunaris	3	49.703	**	-	-	-
S3T1-a							
	Catostomus spp.	3	167.045	**	15	3.541	-
	P. omiscomaycus	3	163.730	**	-	-	-
S3T1-b							
	E. nigrum	3	62.266	**	21	542.721	**
	Catostomus spp.	3	49.512	**	21	295.016	**

Table 4 (continued)

	P. flavescens	3	66.310	**	21	121.055	**
	P. semilunaris	3	74.854	**	21	397.104	**
	P. omiscomaycus	3	979.052	**	-	-	-
S3T1-c							
	E. nigrum	3	1.404	-	17	489.647	**
	P. flavescens	3	7.456	-	17	1515.311	**
	Catostomus spp.	3	23.064	**	17	234.490	**
	M. salmoides	3	20.458	**	17	237.256	**
	P. semilunaris	3	70.314	**	17	286.267	**
	E. lucius	3	27.119	**	17	106.094	**
	N. crysoleucas	3	28.303	**	17	99.116	**
	P. caprodes	3	29.422	**	17	105.692	**
	A. rupestris	3	30.684	**	17	131.479	**
	O. mordax	3	30.429	**	16	132.942	**
	P. omiscomaycus	3	2504.148	**	-	-	-

recovered sequences, threshold values were set, data was filtered and observations were made from the unfiltered and filtered datasets.

EXPERIMENTAL RESULTS

Caveat

Here we present results from several experiments designed to provide insight into the limits of detection for metabarcoding approaches to species level identification. There is, however, a level of uncertainty associated with the lowest limits of detection reported for each target species and for signal intensities produced by all taxa. Our uncertainty relates to the initial tissue used to construct community fish samples and originates from measurement error and replication. Measurement errors occurring during sample construction can alter biomass ratios from expected values and result in skewed sequence diversity. The error associated with tissue mass measurements and homogenate concentrations were ± 0.004 mg and ± 0.00076 mg/uL, respectively. For S1, there was little difference between the initial and error adjusted ratios of target biomass to total sample mass (i.e., tested target detection levels were unaffected by measurement error). Mean error for S2 measurements was ± 0.152 mg/taxa/sample for non-targets and for target when target biomass was ≥ 20% total sample mass; mean error for target was ± 0.0532 mg/sample when target biomass ≤ 0.1%. In S3T1, mean error was ± 0.17 mg/taxa/sample for non-targets and mean error was ± 0.238 mg/sample for target when target biomass was ≥

1% of total sample mass; target mean error was ± 0.010 mg/sample when target biomass was ≤ 0.33%. Although replication for the pilot study S1 was sacrificed in order to increase the range of tested detection levels and fulfill the S1 objective as a range finding experiment, increased replication in subsequent experiments provided insight into the extent of measurement error occurring across sample sets.

Experimental Sample Set 1 (S1), pilot study: Defining an appropriate range of detection levels to test further and initial evaluation of processing based detection errors

Sequencing of the CO1 marker in S1 generated 66,921written sequences assigned expected and unexpected taxonomy after denoising and preprocessing (including chimera isolation) with an average length of 563 bp. Based on S1 PTP layout, the total number of sequences generated per sample was very low,on average 24% less than the minimum expectation based on plating (Appendix B.1). Before filtering there were 59,785 written sequences assigned expected non-target or target species level taxonomy. Marked differences were observed between genetic signals associated with taxa such that the relative frequency of sequences did not correspond to relative biomass abundance and the degree of correspondence varied across taxa. The two measurements varied significantly for all non-targets (Table 4b) across replicates where target biomass was ≤ 10% of total sample mass (N = 13; Fig. 2a). The most notable differences were observed between non-targets *L. lota* and *Catostomus spp.*; at approximately 58% of total sequences, signal for non-target *L. lota* represented a much larger proportion compared to *Catostomus spp.* signal that was considerably under-represented at 2.8% (Fig. 2a). A similar pattern was observed for equal proportion controls where signal strength significantly deviated (Table 4a) from corresponding relative biomass abundance for all non-targets as well as the target *O. mordax* (N = 2; Fig. 2b).

In the unfiltered data set, signal for the S1 target taxon was detected as a single or double sequence hit (i.e., singleton or doubleton) in individual control replicates (N = 2; Fig. 2c). Target signal was also detected in the equal proportion controls with a 1:5 detection level where target biomass was 20% of total sample mass (N = 2; Fig. 2c); here the signal represented 0.57% of total sequences (i.e., almost 97% fewer sequences than expected). Target signal was no longer detected after filtering when approximately 24% of written sequences below threshold values were removed to reduce sequencing noise (N = 15; Fig. 2c). After filtering, overall non-target detection rates also declined due to signal losses sustained by *P. nigromaculatus* and *Catostomus spp.* in samples containing ≥ 10% target biomass. Although the target was undetected after S1 data analysis, further investigation into the legitimacy of the non-detection events resulted in target signal amplification. In total, 2,387 chimeric sequences were discovered to have received similar species level taxonomy for each fragment with minor variations in nomenclature, a flaw in our reference database that resulted in a false non-detection event. Target sequence recovery led to improved detection success rates and expanded range of detection (Fig. 2c) to samples with approximately 1% target biomass; although in this sample, positive detection resulted from a singleton hit. Observations made *after* sequences were recovered, suggest the CO1 fish barcode is detectable when target biomass is ≥ 1% of total sample mass. However, this detection limit remains provisional as detection was not evaluated when target biomass was present at levels between 1 and 0.1% total sample mass (samples with 0.1% target biomass were the first samples where target was not detected).

Collectively, results from S1 led to modifications in the next experimental sample Set 2 (S2) design, methods and approach to data analysis. S1 results provide evidence that non-detection risk is inflated for taxa that are initially present at low to moderate/high abundance, but genetic signal is under-represented. Prior to sequence recovery, weak signal removal resulted in complete non-detection of target signal. Detection rates also declined for two non-targets *Catostomus spp.* and *P. nigromaculatus* with under-represented genetic signals despite uniform non-target biomass abundance. In addition to what was learned about weak signal removal and non-detection risk, S1 results provided insight into the need for reference sequence database revisions to decrease the likelihood for detection errors.

Experimental Sample Set 2 (S2): Modified design to test restricted range of detection levels

Sequencing of CO1 marker in S2 samples generated 718,615 written sequences (over 350,000 per run), with an average length of 633 bp. In contrast to the number of sequences generated per sample for S1, total sequences generated for S2 was on average 29% greater than the minimum expectation (Appendix B.1) based on the S2 PTP layout that was modified from the S1 plating strategy (Appendix B.1). Expected non-target or target taxonomy was assigned to 99.7% of sequences; in total 2.92% of these expected sequences fell below threshold values and were filtered from the final data set. Analogous to S1, considerable differences were observed between total sequences assigned to each taxa; despite PCR primer design modifications aimed to reduce the potential for interspecific PCR bias (Appendix A, Table A.2) genetic signals did not correspond to relative biomass abundances. Significant variation between non-target signals and corresponding biomass abundance (Table 4b) was observed for S2 replicates where target biomass was ≤ 0.1% of total sample mass (N = 13; Fig. 3a). At approximately 54.17% signal for non-target *N. hudsonius* represented a much larger proportion of total sequences, than expected. Conversely, the proportion of sequences assigned to non-targets *G. aculeatus, E. lucius* and *A. rupestris* was much smaller at 19.9%, 14.9% and 13.5% of total sequences, respectively (Fig. 3a). Similarly, the relationship between genetic signal and corresponding relative abundance varied significantly

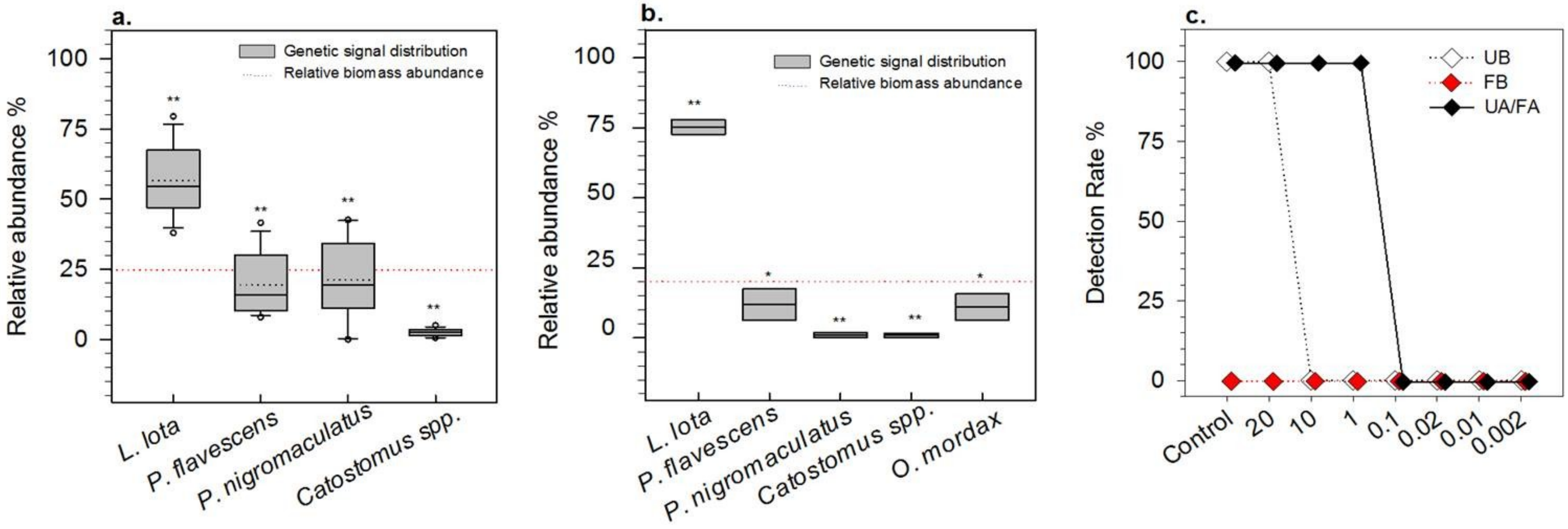


Figure 2 454 sequence data results for experimental sample Set 1 (S1). Samples were constructed using adult fish tissue from four non-target and one target species (*O. mordax*). All samples were comprised of equal non-target biomass proportions and spiked with target tissue to achieve the target detection level (2c, x-axis). **2a.** Observed genetic signal for S1 non-target taxa. Genetic signal is represented by relative sequence abundance as a percent of total sequences (y-axis) generated for non-target taxa in replicates constructed with equal proportions of non-target biomass and when target biomass was < 10% of total sample mass (N = 13). **2b.** Observed genetic signals for S1 non-target and target taxa. Genetic signal is represented by relative sequence abundance as a percent of total sequences generated for non-target and target taxa in control replicates constructed with equal proportions of target and non-target biomass or a 1:5, target to non-target biomass ratio (N = 2). (Fig. 2a and 2b, boxplots show median values (solid horizontal line), mean values (black dotted line), interquartile range (box outline), 90th percentile (whiskers) and outlier values (open circles); red dotted line represents relative biomass abundance per taxa used to construct samples. Groups that were significantly different from relative biomass abundance are indicated by an asterisk (* p < 0.05; ** p < 0.001)). **2c.** Progression of detection

Figure 2 (continued)

success rates for S1 target *O. mordax.* Detection rates (y-axis) represent the percent of replicates of the individual control and spiked experimental mixes where genetic data was generated for S1 target at each target detection level (i.e., target biomass proportion relative to total sample mass (x-axis) in the unfiltered, then in filtered data (i.e. before and after filtering sequence data to reduce the probability of false positive detection errors) both before (UB, FB) and after (UA,FA), respectively, correcting false non-detection errors. Data reflects individual control replicates (N = 2), equal proportion controls, 20% (N = 2) and experimental mixes comprised of 10, 1, 0.1, 0.002% (N = 1), 0.1% (N = 2), 0.02%, 0.01% (N = 3) target biomass (progression of data analysis demonstrating the effect of data processing on positive detection; 1) UB, 2) FB, 3) UA, 4) FA. After correcting detection errors, there was no change between detection rates observed for the unfiltered and filtered datasets).

in equal proportion control replicates (Table 4a) where each taxa comprised 20% of total sample biomass (N = 4; Fig. 3b). Here signal for target *P. semilunaris* comprised a significantly smaller proportion at 4.26% of total sequences, whereas the signal for non-target *N. hudsonius* remained significantly over-represented (Fig. 3b).

Detection success rates for non-target signals were initially 100% for all S2 samples; signals were strong and detection rates did not decline after weak signal removal. In the unfiltered data set, S2 target *P. semilunaris* signal was detected in the control (N = 1) and in equal proportion controls at a rate of 100% (N = 4; Fig. 3c). Target signal was present in 50% and 25% of replicates where target biomass comprised 0.1% (N =4) and 0.02% (N = 4), respectively, of total sample mass (Fig. 3c), but detection success was attributed to only 1 – 3 sequence hits. In replicates with ≤ 0.1% target biomass, target signal fell below threshold values, after filtering target signal was undetected (N = 8; Fig. 3c). Further investigation into the legitimacy of these non-detection events did not result in target signal amplification.

Experimental Sample Set 3 Treatment 1 (S3T1): Assessing non-detection risk associated with increased sample complexity related to species richness

Sequencing of CO1 marker in S3T1 samples constructed with one target *P. omiscomaycus* and one (S3T1-a), four (S3T1-b) or ten (S3T1-c) non-target taxa generated 270,040 written sequences with an average length of 592 bp. Before weak signal removal, 97.3% were assigned expected species level taxonomy. On average 0.25%, 0.53% and 1.08% of total expected sequences were removed from S3T1-a, b, and c, respectively. Ten of eleven S3T1 taxa were positively detected in corresponding individual taxa controls (N = 1 per taxa). Genetic signal for non-target *M. salmoides* was not detected in the control and was only observed in 1of 22 samples expected to contain *M. salmoides* DNA (Fig. 6d). Further investigation into the legitimacy of this non-detection event resulted in signal amplification and positive detection for *M. salmoides* (Fig. 7). Reassigning taxonomy from a larger reference database revealed 4,410 sequences that had initially been identified as Perciformes *spp.* were also *M. salmoides* when identified to a higher taxonomic resolution. Reported results are from unfiltered and filtered data after sequence recovery.

In S3T1-a replicates where target biomass was ≤ 1%, non-target genetic signal did not deviate from corresponding relative biomass abundance (N = 16; Fig. 4a). But, significant variation was observed in replicates with a 1:1 biomass ratio (Table 4a) where signal for target *P. omiscomaycus* represented approximately 95-99% of total sequences (N = 4; Fig. 4b). Target signal was detected at a rate of 100% across experimental mixes (N = 20; Fig. 4c) with the lowest limit occurring in samples comprised of approximately 0.125% (lowest

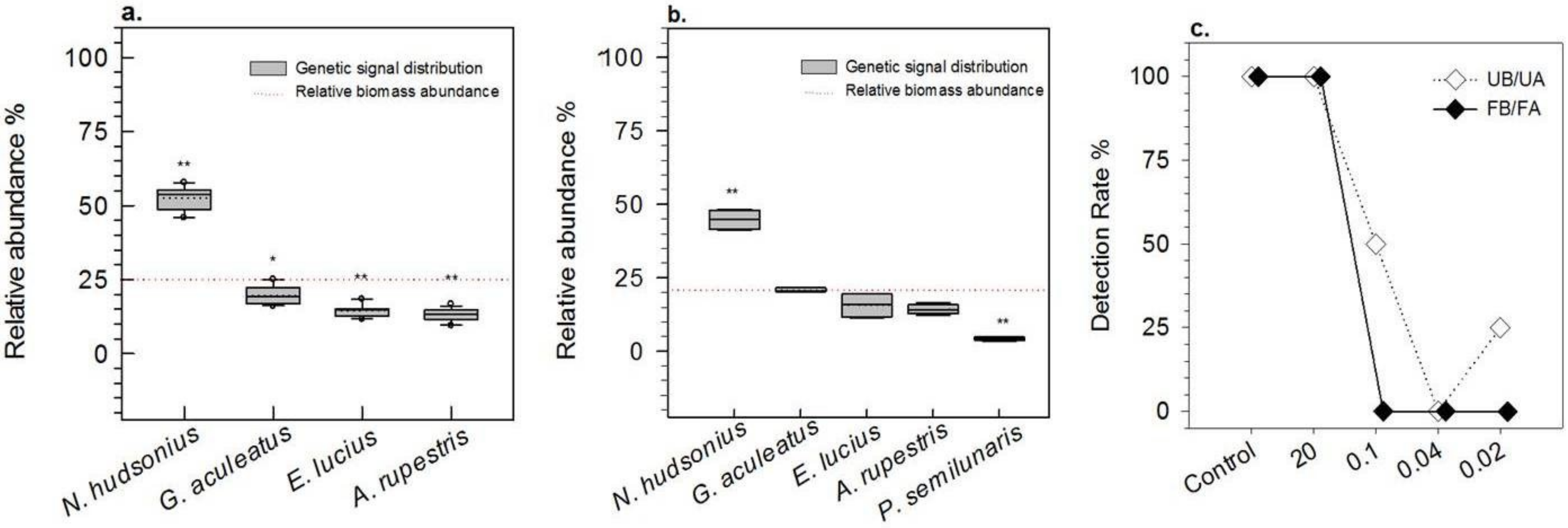


Figure 3 454 sequence data results for experimental sample Set 2 (S2). Samples were constructed using larval fish tissue from four non-target and one target species (*P. semilunaris).* All samples were comprised of equal non-target biomass proportions and spiked with decreasing proportions of target tissue to achieve the target detection level (3c, x-axis). **3a.** Observed genetic signal for S2 non-target taxa. Genetic signal is represented by relative sequence abundance as a percent of total sequences (y-axis) generated for non-target taxa in replicates constructed with equal proportions of non-target biomass and when target biomass was ≤ 0.1 % of total sample mass (N = 13). **3b.** Observed genetic signals for S2 non-target and target taxa. Genetic signal is represented by relative sequence abundance as a percent of total sequences generated for non-target and target taxa in control replicates constructed with equal proportions of target and non-target biomass or a 1:2, target to non-target biomass ratio, (N = 4). (Figure 3a, 3b boxplots show median values (solid horizontal line), mean values (black dotted line) interquartile range (box outline), 90th percentile (whiskers) and outlier values (open circles); red dotted line represents relative biomass abundance per taxa used to

Figure 3 (continued)

construct samples. Groups that were significantly different from relative biomass abundance (expected distribution) are indicated by an asterisk (* p < 0.05; ** p < 0.001)). **2c.** Progression of detection success rates for S2 target *P. semilunaris.* Detection rates (y-axis) represent the percent of replicates of the individual control and spiked experimental mixes where genetic data was generated for S2 target at each target detection level (i.e., target biomass proportion relative to total sample mass (x-axis) in the unfiltered, then in filtered data (i.e. before and after filtering sequence data to reduce the probability of false positive detection errors) both before (UB, FB) and after (UA,FA), respectively, correcting false non-detection errors. Data reflects individual control replicates (N = 1), equal proportion controls, 20% (N = 4) and experimental mixes where N = 4 for each detection level. The progression of data analysis demonstrates the effect of data processing on positive detection; 1) UB/UA, 2) FB/FA. Results from two datasets are reported because each non-detection event appeared legitimate; therefore, there was no difference between UB, UA or FB, FA.

detection level tested for S3T1-a) target biomass; in addition, detection was unaffected after filtering. Signal for S3T1-a non-target *Catostomus spp.* was also detected in 100% of replicates (N = 20) before and after filtering.

Signal for target *P. omiscomaycus* was also significantly over-represented (Table 4a) in S3T1-b equal proportion control replicates (N = 4; Fig. 5b) where target biomass represented 20% of total sample mass and non-target signals were significantly less than corresponding relative biomass abundance (Table 4a) (N = 4; Fig. 5b). Signals produced by non-target taxa also varied significantly from expected values (Table 4b) in replicates with ≤ 1% target biomass (N = 20; Fig. 5a). Target signal was detected in 100% of experimental mixes (N = 24; Fig. 5c) with the LLD occurring in sample comprised of approximately 0.05% target biomass (lowest detection level tested for S3T1-b). Non-target signals were also observed in 100% of S3T1-b samples (N = 26). Weak signal removal did not affect detection of S3T1-b taxa.

Significant over-representation of target signal (Table 4a) persisted in S3T1-c equal proportion control replicates (N = 4; Fig. 6b) and for some non-target taxa, genetic signal strength was significantly less (Table 4a) than corresponding relative biomass abundance (N = 4; Fig. 6b). Non-target signals also varied significantly from expected values (Table 4b) in replicates with ≤ 1% target biomass (N = 16; Fig. 6a). As was observed in S3T1 samples constructed with < 10 non-target taxa (i.e., S3T1-a, b), target signal was detected in all

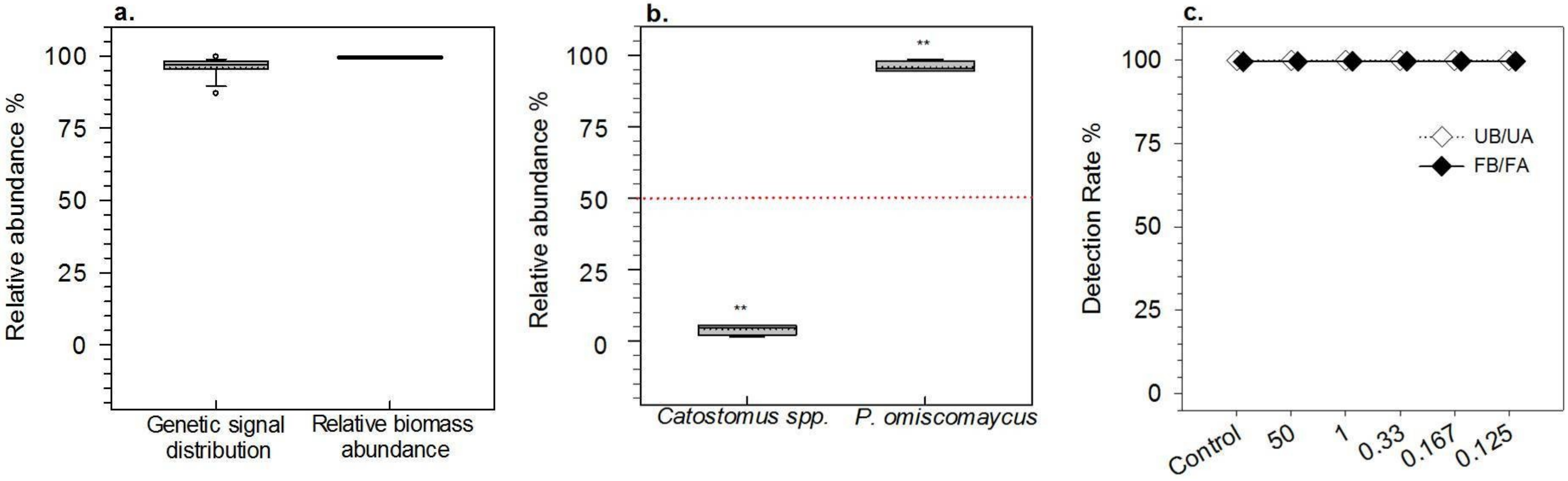


Figure 4 454 sequence data results for experimental sample Set 3 Treatment 1 (S3T1) sub-set a (S3T1-a). S3T1-a was part of a larger sample set that included samples constructed with five (Figure 5, S3T1-b) and eleven (Figure 6, S3T1-c) total taxa which was designed to assess the effect of increasing species richness on positive detection. S3T1-a was constructed using larval fish tissue from one non-target and one target species (*P. omiscomaycus*). Equal proportion controls were comprised of equal non-target and target biomass (N = 4) and experimental mixes were spiked with decreasing proportions of target biomass to achieve the target detection level (4c, x-axis). **4a.** Observed genetic signal for S3T1-a non-target species. Genetic signal is represented by relative sequence abundance as a percent of total sequences (y-axis) for the non-target species in replicates where target biomass was < 50 % (N = 16). **4b.** Observed genetic signals for S3T1-a non-target and target taxa. Genetic signal is represented by relative sequence abundance as a percent of total sequences for each taxa in control replicates constructed with equal proportions of target and non-target biomass or a 1:5, target to non-target biomass ratio, (N = 4). (Figure 4a, 4b boxplots show median values (solid horizontal line), mean values (black dotted line), interquartile range (box outline), 90th percentile (whiskers) and outlier values (open circles);

Figure 4 (continued)

red dotted line represents relative biomass abundance per taxa used to construct samples. Groups that were significantly different from relative biomass abundance (expected distribution) are indicated by an asterisk (* p < 0.05; ** p < 0.001)). **4c.** Progression of detection success rates for S3T1-a target *P. omiscomaycus.* Detection rates (y-axis) represent the percent of replicates for the individual control and spiked experimental mixes where genetic data was generated for S3T1-a target at each target detection level (i.e., target biomass proportion relative to total sample mass (x-axis) in the unfiltered, then in filtered data (i.e. before and after filtering sequence data to reduce the probability of false positive detection errors) both before (UB, FB) and after (UA,FA), respectively, correcting false non-detection errors. Data reflects individual control replicates (N = 1), equal proportion controls, 50% target biomass (N = 4) and experimental mixes where N = 4 for each detection level. The progression of data analysis demonstrates the effect of data processing on positive detection: 1) UB/UA, 2) FB/FA. Results from two datasets are reported because S3T1-a target was detected in 100% of replicates.

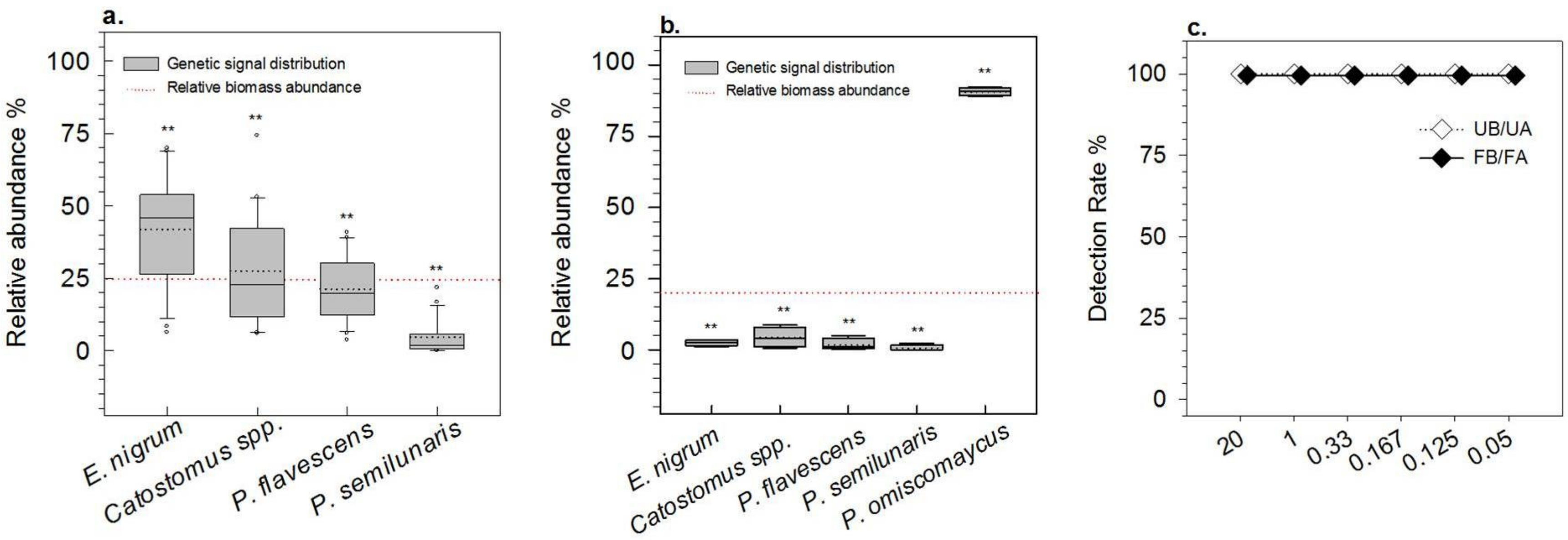


Figure 5 454 sequence data results for experimental sample Set 3 Treatment 1 (S3T1) sub-set b (S3T1-b). S3T1-b was part of a larger sample set that included samples constructed with two (Figure 4, S3T1-a) and eleven (Figure 6, S3T1-c) total taxa which was designed to assess the effect of increasing species richness on positive detection. S3T1-b was constructed using larval fish tissue from four non-target and one target species (*P. omiscomaycus*). Equal proportion controls were comprised of equal non-target and target biomass (N = 4) and experimental mixes were spiked with decreasing proportions of target biomass to achieve the target detection level (5c, x-axis). **5a.** Observed genetic signal for S3T1-b non-target taxa. Genetic signal is represented by relative sequence abundance as a percent of total sequences (y-axis) for non-target taxa in replicates constructed with equal proportions of non-target biomass and where target biomass was < 20% (N = 22). **5b.** Observed genetic signals for S3T1-b non-target and target taxa. Genetic signal is represented by relative sequence abundance as a percent of total sequences for each taxa in control replicates constructed with equal proportions of target and non-target biomass or a 1:5, target to non-target biomass ratio, (N = 4).

Figure 5 (continued)

(Figure 5a, 5b boxplots show median values (solid horizontal line), mean values (black dotted line), interquartile range (box outline), 90th percentile (whiskers) and outlier values (open circles); red dotted line represents relative biomass abundance per taxa used to construct samples. Groups that were significantly different from relative biomass abundance (expected distribution) are indicated by an asterisk (* $p < 0.05$; ** $p < 0.001$)). **5c.** Progression of detection success rates for S3T1-b target *P. omiscomaycus.* Detection rates (y-axis) represent the percent of replicates for spiked experimental mixes where genetic data was generated for S3T1-b target at each target detection level (i.e., target biomass proportion relative to total sample mass (x-axis) in the unfiltered, then in filtered data (i.e. before and after filtering sequence data to reduce the probability of false positive detection errors) both before (UB, FB) and after (UA,FA), respectively, correcting false non-detection errors. Data reflects equal proportion controls, 20% target biomass (N = 4) and experimental mixes where N = 4 for each detection level. The progression of data analysis demonstrates the effect of data processing on positive detection: 1) UB/UA, 2) FB/FA. Results from two datasets are reported because S3T1-b target was detected in 100% of replicates.

experimental mixes at a rate of 100% with the LLD occurring in samples comprised of approximately 0.125% target biomass (lowest detection level tested for S3T1-c) and detection rates did not change after weak signal removal (N = 20; Fig. 6c). In contrast to S3T1-a and b detection rates for S3T1-c non-target taxa varied interspecifically and rates declined for some non-targets after filtering sequences below threshold values (N = 20; Fig. 7). *Experimental Sample Set 3 Treatment 2 (S3T2) pilot study: Assessing non-detection risk associated with field sample processing efforts*

The initial round of PCR failed to generate CO1 amplicons and samples were not sequenced. Investigation into the failed PCR (Appendix B.2) led to method modifications and a new S3T2 design. Despite modifications, PCR did not work for controls comprised of detritus homogenate (1:0 detritus to fish ratio) nor for a replicate of the experimental mix with a 1:1 ratio; these samples were not sequenced. Pyrosequencing of CO1 markers generated 17,038 written sequences with expected species level taxonomy for S3T2 individual controls (N = 5), larval fish mix controls (1:5, N = 3; 1:100, N = 3) and experimental mixes (N = 19). In total, 2.67% of these sequences fell below set threshold values and were filtered from the final data set.

Taxa specific effects and detrital effects on positive detection were observed in S3T2 replicates. Signal produced by target *P. omiscomaycus* was

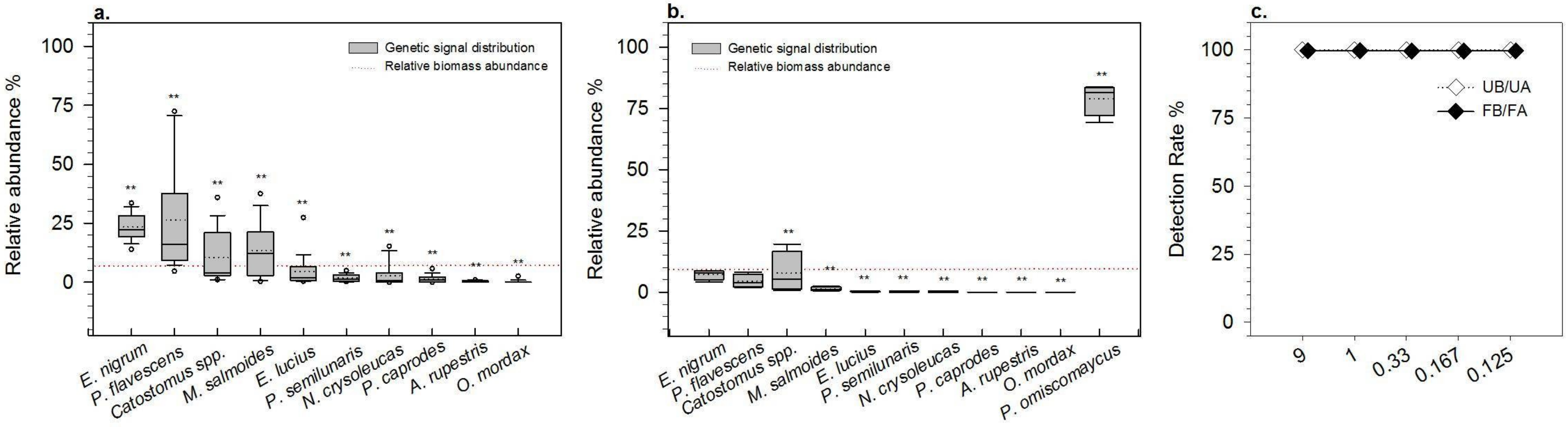


Figure 6 454 sequence data results for experimental sample Set 3 Treatment 1 (S3T1) sub-set c (S3T1-c). S3T1-c was part of a larger sample set that included samples constructed with two (Figure 4, S3T1-a) and five (Figure 5, S3T1-b) total taxa which was designed to assess the effect of increasing species richness on positive detection. S3T1-c was constructed using larval fish tissue from ten non-target and one target species (*P. omiscomaycus*). Equal proportion controls were comprised of equal amounts of biomass for each non-target and target (N = 4) and experimental mixes were spiked with decreasing proportions of target biomass to achieve the target detection level (6c, x-axis). **6a.** Observed genetic signal for S3T1-c non-target taxa. Genetic signal is represented by relative sequence abundance as a percent of total sequences (y-axis) for non-target taxa in replicates constructed with equal proportions of non-target biomass and where target biomass was < 9% (N = 18). **6b.** Observed genetic signals for S3T1-b non-target and target taxa. Genetic signal is represented by relative sequence abundance as a percent of total sequences for each taxa in control replicates constructed with equal proportions of target and non-target biomass or a 1:11, target to non-target biomass ratio, (N = 4). (Figure 6a, 6b boxplots show median values (solid horizontal line), mean values (black dotted line), interquartile range (box outline), 90th percentile (whiskers) and outlier values (open circles); red dotted line represents relative biomass abundance per taxa used to construct samples. Groups

Figure 6 (continued)

that were significantly different from relative biomass abundance (expected distribution) are indicated by an asterisk (* $p < 0.05$; ** $p < 0.001$)). **6c.** Progression of detection success rates for S3T1-b target *P. omiscomaycus.* Detection rates (y-axis) represent the percent of replicates for spiked experimental mixes where genetic data was generated for S3T1-c target at each target detection level (i.e., target biomass proportion relative to total sample mass (x-axis) in the unfiltered, then in filtered data (i.e. before and after filtering sequence data to reduce the probability of false positive detection errors) both before (UB, FB) and after (UA,FA), respectively, correcting false non-detection errors. Data reflects equal proportion controls, 9% target biomass (N = 4) and experimental mixes where N = 4 for each detection level. The progression of data analysis demonstrates the effect of data processing on positive detection: 1) UB/UA, 2) FB/FA. Results from two datasets are reported because S3T1-b target was detected in 100% of replicates.

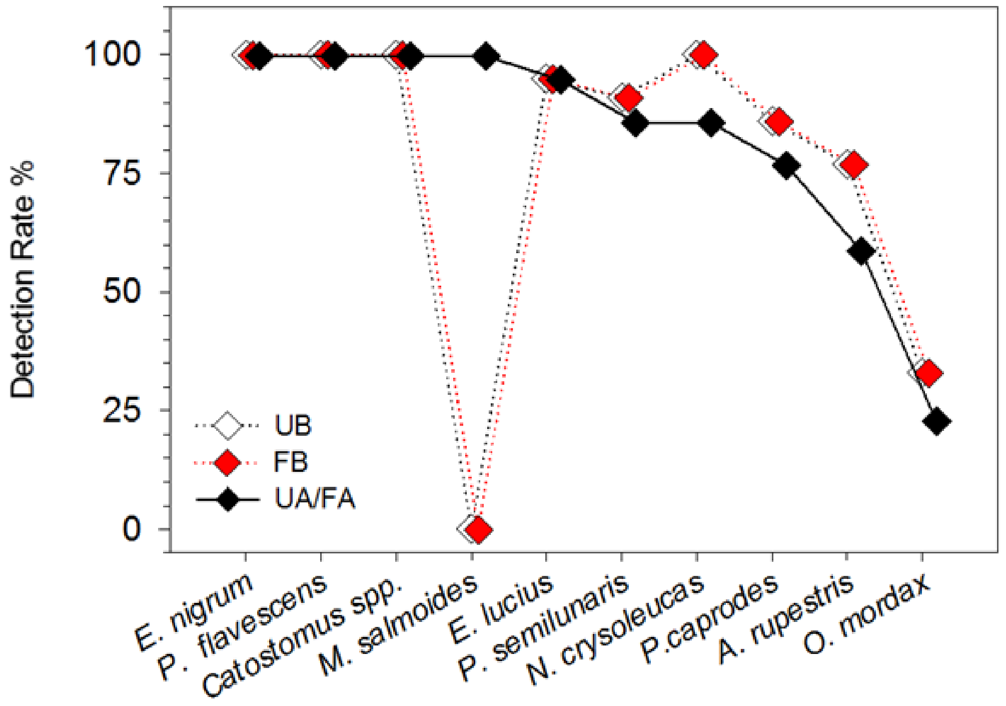


Figure 7 Progression of detection success rates derived from 454 sequence data results for experimental sample Set 3 Treatment 1 (S3T1) sub-set c (S3T1-c) non-target taxa. S3T1-c was part of a larger sample set that included samples constructed with two (Figure 4, S3T1-a) and five (Figure 5, S3T1-b) total taxa which was designed to assess the effect of increasing species richness on positive detection. S3T1-c was constructed using larval fish tissue from ten non-target and one target species (*P. omiscomaycus*). Detection rates (y-axis) represent the percent of samples out of total samples (N = 18) where genetic data was generated for each non-target species in the unfiltered, then in filtered data (i.e. before and after filtering sequence data to reduce the probability of false positive detection errors) both before (UB, FB) and after (UA,FA), respectively, correcting false non-detection errors. Each non-target comprised approximately 9.9% of total sample mass. The progression of data analysis demonstrates the effect of data processing on positive detection: 1) UB/UA, 2) FB/FA. After correcting detection errors there was no change between detection rates observed for the unfiltered and filtered datasets).

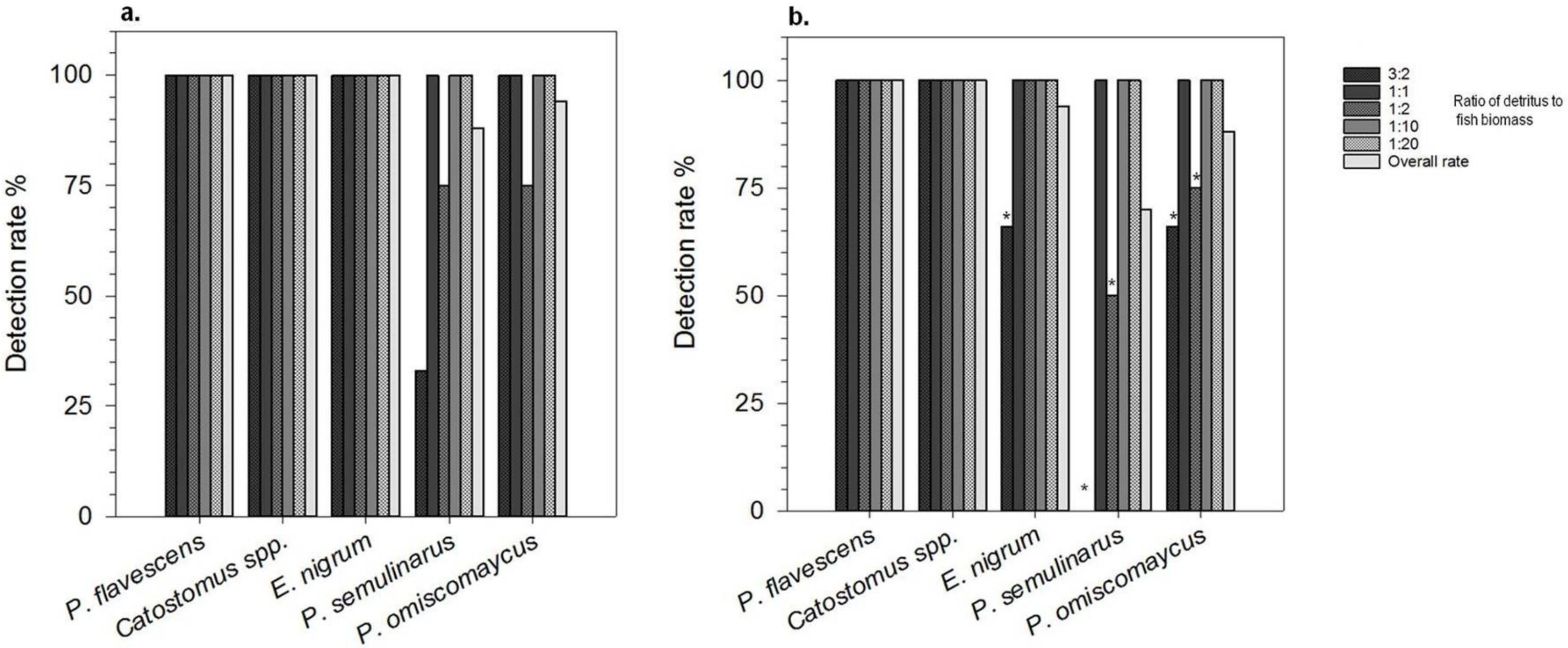


Figure 8 Progression of detection success rates derived from 454 sequence data results for experimental sample Set 3 Treatment 2 (S3T2) evaluating the effect of detritus presence on detection success. Samples were constructed with larval fish tissue from four non-target and one target (*P.* omiscomaycus) taxa. Experimental mixes were comprised of a detritus to fish biomass ratio (figure legend) with equal proportions of non-target biomass spiked with target biomass to achieve a target to non-target biomass ratio of 1:100. Detection success rates (y-axis) are the percent of replicates for each condition (detritus to fish biomass ratio) where genetic data was generated for non-target and target taxa in each detritus condition in the **a.** unfiltered sequence data and for **b.** filtered data, target detection level (for each taxa, N = 19) asterisk indicates changes in detection rates after filtering.

over-represented in the equal proportion fish controls (N = 3) with no detritus and 20% biomass per taxa. Non-target genetic signals did not correspond to relative biomass abundance in fish biomass control replicates or experimental mixes where target biomass represented 1% of total sample mass. A general pattern was observed where non-targets *P. flavescens* and *Catostomus spp.* were consistently over-represented whereas non-targets *E. nigrum* and *P. semilunaris* were consistently under-represented, with *P. semilunaris* being the more extreme (N = 22) except in equal proportion controls (N = 3). Before filtering for weak signals, only *P. semilunaris* and target *P. omiscomaycus* were detected in less than 100% of replicates with non-detection events only occurring in experimental mixes (N = 19; Fig. 8a). After filtering, the overall detection rate declined for non-targets *E. nigrum* and *P. semilunaris* and for target *P. omiscomaycus* (3:2; N = 3; Fig. 8b). Lower detection rates and rate declines following filtering were mainly associated with signal loss from samples constructed with the greatest detritus to fish biomass ratio at 3:2 (N = 3).

DISCUSSION

Detection sensitivity afforded by metabarcoding analysis of fish communities

High-throughput sequencing technology may provide a rapid, cost effective method for species richness determination in complex community samples (Ji *et al.* 2013; Ko *et al.* 2013; Stein *et al.* 2014). Given the costs associated with traditional taxonomic identification of many aquatic organisms, metabarcoding analyses have gained recognition as potentially powerful tools for early detection of aquatic invasive species. A practical early detection strategy, however, demands balancing detection costs with an acceptable level of non-detection risk (Lodge *et al.* 2006; Trebitz *et al.* 2009; Hoffman *et al.* 2011). Although a few recent studies have empirically shown metabarcoding to be accurate and highly sensitive regarding detection in some aquatic invertebrate communities (Hajibabaei *et al.* 2011; Pochon *et al.* 2013; Zhan *et al.* 2013) the limits to detection in multi-species assemblages for fish communities has not been reported.

Here we evaluated non-detection risk associated with some standard metabarcoding methods by constructing artificial community samples with known species richness and relative biomass abundance composed of tissue from multiple “non-target” fish taxa and spiked with various proportions “target” fish tissue from a single species not already present in the sample. Our main findings provided convincing experimental evidence that we can detect genetic signals produced by spiked target species comprising as low as 0.02% - 1% of the original total sample biomass (Fig. 2c - 6c), and demonstrated that the lowest limit of detection (LLD) observed for each target species varied between experiments. Our ability to detect and the associated risk of non-detection, regardless of starting biomass, appeared susceptible to several factors that skewed genetic signals from corresponding biomass abundance including CO1 amplification bias, sequence data processing methods and reference sequence database composition. On the basis of sample complexity, S3T1 results suggest increasing species richness does not impede our ability to detect this target within these sample matrices (Fig. 4c – 6c), but impacts of species richness on target detection may be confounded by PCR bias exhibited by the specific mix of taxa. Furthermore, as a tentative set, results from S3T2 demonstrate the inclusion of detrital material can inhibit our ability to detect and determine full species richness in some samples (Fig. 8). Here we discuss our theoretical expectations for detection sensitivity levels associated with standard metabarcoding methods and examine our findings in context of the body of experiments to illustrate the methodological and analytical factors influencing the risk of non-detection. Finally, we highlight our main conclusions and discuss the next research steps for applying this tool to early detection of aquatic invasive species.

The range of detection sensitivity levels determined from our experiments coincides with detection sensitivities previously reported for aquatic communities (Hajibabaei *et al.* 2011; Pochon *et al.* 2013). At 1% of total sample biomass, the LLD for S1 (Fig. 4c) was similar to the LLD described in Hajibabaei *et al*. (2011). However, becauseS1 had roughly 5X fewer interacting species than described in Hajibabaei *et al.* (2011) and in a simple matrix with four taxa we anticipated positive detection of target signal in samples with much lower probabilities of detection. Through S2 experimental design and method modifications, we expected the LLD to exceed S1 and the LLD for S2 target at 0.02% of total sample mass was 50X greater than the limit observed for S1 (Fig. 3c). This result suggests the modifications improved our ability to recover sequences for low abundance taxa in context of the S2 experiment and in general, the LLD for fish taxa in a simple sample matrix is lower than we initially had achieved in S1.

Methodological and analytical factors influencing the risk of non-detection

Collectively, results from these experimental sets demonstrate our ability to detect taxa in multi-species samples, regardless of starting biomass abundance (i.e., common or rare taxa), can be influenced by several factors. Experimentally, the absence of target genetic signal can indicate a limit to detection, but the lack of some non-target signals, despite having ample starting material (i.e., tissue homogenate), suggests uncontrolled factors may be inflating the risk of non-detection and leading to detection errors. For S1 and S2 targets, we may explain signal absence as the limit of detection primarily because detection at the lowest limit was associated with a single sequence hit and detection was not replicated. However, the observed genetic signal skew in unison with reduced detection rates suggest the risk of non-detection was inflated in some cases. Correspondingly, genetic signal skew occurring in S3T1-c (Fig. 6a, b) appeared to increase non-detection risk for approximately 50% of non-target taxa. We did not detect signals for some non-targets even though non-target tissue was abundant relative to target tissue; furthermore, when we were able to recover sequences for these taxa the observed signals were usually represented by very few sequences. Reduced signal strength and signal absence, despite sufficient starting tissue material, may result from measurement error occurring within and between samples. Although there may have been some degree of measurement error, it is unlikely the error was large enough to produce the observed signal differences. Moreover, the signal skew was similar between replicates for each experiment as was the cross-set signal pattern observed for *P. semilunaris*. In addition, signal distribution across replicates was mostly confined to a small range of values (Fig. 2 – 6, a, b) with the wider distributions associated probably resulting from slight differences between sample composition for each tested detection level and/or measurement error. Therefore, it was likely PCR bias and not measurement error caused the observed genetic signal skew and thereby increased non-detection risk for some taxa.

PCR bias may have caused the observed signal skew in each sample but multiple sources of bias have been reported including interspecific variation in gene copy number, denaturation efficiency, primer binding affinity and PCR drift (i.e., random amplification) (Wagner *et al.* 1994; Polz & Cavanaugh 1998; Ishii & Fukui 2001). Although the exact source for bias in our samples in unknown, the base composition of the CO1 barcode region, our PCR design and primer(s) can help narrow the possibilities. Interspecific variation in gene copy number is part of the inherent variation observed in field samples. Mitochondria densities in fishes can vary between tissue types (Urschel & O'Brien 2008) within individuals, species and between species (Bennett & Johnston 2008); in addition, density differences can depend on life stage (Bennett & Johnston 2008) and on water temperature (Guderley & Johnston 1996; Hardewig *et al.* 1999; Guderley & St-Pierre 2002; Bennett & Johnston 2008). Therefore, as part of the mitochondrial genome, there also could be large differences in CO1 densities. Although the extent in larval fish communities is unknown, we assumed there was some degree of variation and sought to limit it through specimen quantity and life stage, as well as through tissue preparation methods. Comparisons between sequence data generated from our samples constructed with fish tissue and similar samples constructed from equimolar concentrations of fish DNA extracted from adult fin clips (results not reported here) suggest differential CO1 copy number is not a significant source of bias under these circumstances, as signal patterns were similar between the two sample types.

The remaining sources of bias, PCR drift, denaturation efficiency and primer binding affinity are artifacts of PCR. As our experimental design tried to limit bias due to differential CO1 densities, the PCR cycle used throughout our study was designed to limit bias originating from other sources as well. We attempted to reduce bias resulting from random amplification, a minimal contributor to bias, as reported by Polz *et al.* (1998), by pooling multiple PCR replicates. If drift were the sole cause of bias in our samples, the similar signal skew observed across replicates would not have occurred. Therefore, bias in our constructed fish communities is likely due to a different source, such as differential denaturation efficiency or primer-binding affinity corresponding to the base composition of the target gene.

Bias associated with differential denaturation efficiency or primer-binding affinity corresponds to the base composition of the target gene. Genes with higher guanine/cytosine (GC) ratios (i.e., GC content) require higher denaturation temperatures. This is because the triple hydrogen bond binding the GC base pair has higher thermostability than the double bonded adenine/thymine (AT) pair. If the denaturation temperature is not high enough, DNA from taxa with a GC rich target gene (i.e., GC %, ≥ 50%) will denature more slowly than AT rich genes. Therefore, failure to extend the initial denaturation step of the PCR cycle for samples containing multi-template DNA with a mix of GC and AT rich genes could result in incomplete denaturation and subsequent under-amplification of the GC rich genes (Ishii & Fukui 2001). In contrast, but also because triple H-bonds are more stable, genes with high GC content at primer binding sites have an increased primer binding affinity relative to AT rich sites and when universal primers are used, exhibit favorable bias resulting in over-amplification (Polz and Cavanaugh 1998). To limit bias associated with this source, it has been suggested to use annealing temperatures as low as 45° C for multi-templates containing a mix of AT and GC rich genes and for larger template volumes the lower temperature should be combined with longer annealing time (Ishii & Fukui 2001) since temperature change occurs more slowly in larger volumes. If fish taxa used to construct our samples had GC contents above and below the reported average of 47% (Ward *et al.* 2005) we would expect to see differential CO1 amplification due to variation in denaturation efficiency and primer binding affinity. Our PCR cycle was designed to limit bias associated with differential primer binding energies andincluded a reduced annealing temperature and small template volume of to. Nevertheless biases were observed.

Variation in overall GC content could explain the observed bias, since the initial denaturation step was not extended to allow denaturation of genes with higher thermostability. Yet CO1 is highly conserved and interspecific variation is attributed primarliy to synonomous base substitutions (i.e., neutral mutations) at the third base positions, so we would expect overall GC content to be comparable between taxa and little interspecific variation between denaturation efficiences. Differential GC content of the primer binding site is a plausible explanation for observed bias between taxa. Even though our attempt to reduce bias by modifying the PCR design from a single primer set (S1) to a primer cocktail (S2, S3T1,S3T2) seemed unsuccessful, comparisons between taxa common to S1 and S3T1, S3T2 showed different signal strengths, which may indicate the primer cocktail altered amplification bias for those taxa. For example, in S1 genetic signal was skewed away from non-target *Catostomus spp.* (Fig. 2a, b) and this species was detected in < 100% of samples; whereas in S3T1 and S3T2 signal was neutral or skewed towards *Catostomus spp.* and detection rates were 100%. This signal difference could be attributed to the species-specific response to each primer design, effects of PCR bias in unique sample matirices (community specific effects) or a combination of these factors.

The general pattern that emerged from cross-set comparisons suggests the risk of non-detection is lower for taxa exhibiting favorable PCR bias because barcodes for these taxa are over-amplified, which translates into a stronger genetic signal (i.e., more associated barcode sequences) relative to neutral or under-amplified taxa. Moreover, each experiment showed under-amplified taxa are represented by fewer sequences, which are more likely to be removed when data is filtered to reduce the probability of detection errors associated with weak signals. In our experiments, filtering resulted in false non-detection events for non-target or target taxa associated with few sequences in each sample set. Therefore, weak signal removal can inflate non-detection risk for under-represented taxa that result from a corresponding low relative abundance of starting biomass (i.e., rare taxa) or an under-amplified barcode. Recent studies confer additional support for this conclusion (Zhan 2013, 2014); however, a generalized approach to handling weak signals in the context of the rare species detection has yet to be developed.

In addition to PCR bias and weak signal removal, we learned the reference sequence database used to assign taxonomy to unknown barcode sequences could contain flaws that also lead to false non-detection events. Detection errors occurred for S1 target *O. mordax* and S3T1-c non-target *M. salmoides,* but were corrected after the source of error was discovered. Although

S3T1-c detection errors were not linked to the target, we recognized the ability to detect any species, regardless of relative biomass abundance, is limited by the reference database used to assign taxonomy. Detection errors associated with this factor are easily corrected and database optimization to eliminate the associated non-detection risk is crucial for a practical AIS early detection strategy.

Other workflow components can also affect signal strength and thereby influence positive detection. Although each component was not independently addressed within our study, we identified some areas in need of more work in order to customize parameters and processes to minimize non-detection risk. For example, sample collection, handling, processing and preservation methods, should minimize the chance for DNA degradation and contamination to produce samples that yield high quality DNA. Currently we have the tools and knowledge for the effective collection and preservation of larval fish samples; however, we must ensure these methods are implemented in the field and lab. In addition, the PTP layout (Appendix B.1) and parameter settings for bioinformatics processing should be optimized for detection assays to minimize non-detection risk.

Non-detection risk associated with increased sample complexity relating to species richness and field sample processing efforts

Because of sample complexity arising from variation inherent to collected field samples, we suspected that non-detection risk would increase in more complex samples (i.e., increase with species richness and inclusion of detritus). Detection rates observed for S3T1 with increasing species richness suggested non-detection risk for the target taxon was not inflated when more species were present. Genetic signal, however, was skewed in favor of S3T1 target in equal proportion controls for each richness sub-set (Fig. 4b, 5b, 6b); on average, target signal strength was between 2 and 50 times higher than expected. In this case the risk of non-detection for low abundance taxa was likely reduced by the favorable PCR bias exhibited by S3T1 target. Therefore, the impact of increasing species richness on detection could be confounded and these results are considered more equivocal until we can replicate them using other fish species. Similarly, some of S3T2 results were slightly different from what was expected. In addition to expecting an increased level of non-detection risk we predicted genetic signal skew from S3T2 equal proportion controls to mirror the skew observed in S3T1 replicates with identical composition, but the skew differed and unlike S3T1, S3T2 signals were extremely varied between species. Consequently, we do not have full confidence results reported for S3T2 because the observed variation could arise from detritus inclusion (i.e., a detrital effect), but it could also be the product of measurement errors resulting from modifications made to tissue preparation methods (Appendix B.2). Nevertheless, although tentative, the results (Fig. 8) suggest there may be a limit to the amount of detritus allowed in samples intended for metabarcoding analysis, therefore, some level of sample processing (e.g., separating fish from sample residue) may be necessary in order to successfully sequence community samples. Additional testing is necessary to replicate these results to understand the sources of variation and causes of non-detection associated with S3T2 findings.

Main conclusions

This study has helped us gain insight into the amount of non-detection risk associated with metabarcoding analysis. Overall, our findings suggest that we can detect fish taxa when biomass abundance is low relative to other species in the sample. The lowest limit of detection observed in our experiments was between 0.02% and 1% of total sample mass; however, results indicated that PCR bias can skew genetic signals and inflate non-detection risk, therefore, the limits to detection for metabarcoding analysis of larval fish communities seem to be specific to each species and possibly to community composition. Furthermore, favorable PCR bias appeared to enhance our ability to detect low abundance taxa, whereas an unfavorable bias can reduce that ability, regardless of starting relative biomass abundance. Since we do not yet fully understand PCR bias in fish assemblages, we should proceed with caution when interpreting results from samples with unknown species composition. In addition to PCR bias, weak signal removal and reference sequence database composition also affected detection success. Weak signal removal can result in false non-detection events for under-represented taxa, whether under-representation is due to rarity of biomass in the original sample composition or an unfavorable PCR bias. Detection errors may also arise during standard bioinformatics processes involving taxonomic assignment to unknown sequences or sequence portions if low-resolution taxonomy or species level nomenclature variations occur in the reference sequence database for any sampled taxa.

Future research and implications

We now have a better understanding of the standard workflow processes influence on detection in fish communities, but there remains a need to examine these processes and identify ways we can reduce false non-detection events. Although interspecific variations in primer binding affinities seem to be the most likely source of bias, more work is required to verify this assumption, to quantify the risk of non-detection associated with PCR bias and to determine if PCR design modifications to the PCR cycle, new primers, or different primer combinations can reduce bias and essentially the risk of non-detection. Regarding weak signal removal, we are only beginning to understand its effects on detection; nonetheless, we know this process must be optimized for its intended purpose as well as to reduce the risk of non-detection to an acceptable level. We also know our ability to detect is limited by the reference database, so we can begin revise, build and strengthen our database to improve it for detection assays. We should also examine other aspects of the workflow, such as workflow procedures preceding PCR amplification, as well as parameters used in bioinformatics analyses so to identify the ideal settings. In addition, we need to compare results from detection assays using standard and optimized workflow processes to continue the investigation into non-detection risk associated with the inherent variation observed in field-collected samples. A thorough understanding of the risk of non-detection associated with sample composition and each workflow component (before and after optimizing the standard workflow) should help us determine an acceptable level of non-detection risk. This knowledge is critical to determine if high-throughput metabarcoding analyses can replace traditional taxonomy and balance the associated costs with risk of non-detection in a practical aquatic invasive species early detection strategy.

References

454LifeSciences (2009) Genome Sequencer FLX Titanium Research Applications Guide. 454 Life Science Corp., Branford, CT.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *Journal of molecular biology,* **215,** 403-410.

Arena, A. (2010) Dna exitus plus™ versus standard bleach solution for the removal of dna contaminants on work surfaces and tools. *Investigative Sciences Journal,* **2**.

Auer, N.A. (1982) Identification of larval fishes of the Great Lakes basin with emphasis onthe Lake Michigan drainage. *Special Publication*. Great Lakes Fisheries Commission, Ann Arbor, MI 48105.

Bennett, A.F. & Johnston, I.A. (2008) *Animals and temperature : phenotypic and evolutionary adaptation*. Cambridge University Press, Cambridge [England]; New York.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Wheeler, D.L. (2005) GenBank. *Nucleic Acids Res,* **33,** D34-D38.

Boileau, M. (1985) The expansion of white perch, Morone americana, in the lower Great Lakes. *Fisheries,* **10,** 6-10.

Bott, N.J., Ophel-Keller, K.M., Sierp, M.T., Rowling, K.P., McKay, A.C., Loo, M.G., Tanner, J.E. & Deveney, M.R. (2010) Toward routine, DNA-based detection methods for marine pests. *Biotechnology advances,* **28,** 706-714.

Burden, D.W. (2012) Guide to the Disruption of Biological Samples. *Random Primers, , Jan. 2012, Page 1-25 (updated June 4, 2012)***,** 1-25.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J. & Knight, R. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods,* **7,** 335-336.

CFR (1993) Ballast Water management for Vessels Entering the Great Lakes *Code of Federal Regulations* (ed. U.S.C. Guard).

Crowder, L.B. (1980) Alewife, rainbow smelt and native fishes in Lake Michigan: competition or predation? *Environmental Biology of Fishes,* **5,** 225-233.

Edgar, R.C. (2005) UCLUST user guide.

Folmer, Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology,* **3** 294-299.

French III, J.R. & Edsall, T.A. (1992) Morphology of ruffe (Gymnocephalus cernuus) protolarvae from the St. Louis River, Lake Superior. *Journal of Freshwater Ecology,* **7,** 59-68.

Goldberg, C.S., Sepulveda, A., Ray, A., Baumgardt, J. & Waits, L.P. (2013) Environmental DNA as a new method for early detection of New Zealand mudsnails (Potamopyrgus antipodarum). *Freshwater Science,* **32,** 792-800.

Guderley, H. & Johnston, I. (1996) Plasticity of fish muscle mitochondria with thermal acclimation. *J Exp Biol,* **199,** 1311-1317.

Guderley, H. & St-Pierre, J. (2002) Going with the flow or life in the fast lane: contrasting mitochondrial responses to thermal change. *Journal of Experimental Biology,* **205,** 2237-2249.

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

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A. & Baird, D.J. (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One,* **6,** e17497.

Hall, S. & Mills, E. (2000) Exotic species in large lakes of the world. *Aquatic Ecosystem Health & Management,* **3,** 105-135.

Hardewig, I., Van Dijk, P., Moyes, C. & Pörtner, H.-O. (1999) Temperature-dependent expression of cytochrome-c oxidase in Antarctic and temperate fish. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology,* **277,** R508-R516.

Hebert, P.D., Cywinska, A., Ball, S.L. & deWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proc Biol Sci,* **270,** 313-321.

Hebert, P.D., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proc Natl Acad Sci U S A,* **101,** 14812-14817.

Hebert, P.D., Ratnasingham, S. & deWaard, J.R. (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc Biol Sci,* **270 Suppl 1,** S96-99.

Hebert, P.D., Stoeckle, M.Y., Zemlak, T.S. & Francis, C.M. (2004b) Identification of birds through DNA barcodes. *PLoS Biol,* **2,** e312.

Hecky, R., Smith, R.E., Barton, D., Guildford, S., Taylor, W., Charlton, M. & Howell, T. (2004) The nearshore phosphorus shunt: a consequence of ecosystem engineering by dreissenids in the Laurentian Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences,* **61,** 1285-1293.

Hoffman, J.C., Kelly, J.R., Trebitz, A.S., Peterson, G.S., West, C.W. & Jackson, D. (2011) Effort and potential efficiencies for aquatic non-native species early detection. *Canadian Journal of Fisheries and Aquatic Sciences,* **68,** 2064-2079.

Hulme, P.E. (2006) Beyond control: wider implications for the management of biological invasions. *Journal of Applied Ecology,* **43,** 835-847.

Ishii, K. & Fukui, M. (2001) Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. *Appl Environ Microbiol,* **67,** 3753-3755.

Ivanova, N.V., Zemlak, T.S., Hanner, R.H. & Hebert, P.D.N. (2007) Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes,* **7,** 544-548.

Jackson, J., Laikre, L., Baker, C.S. & Kendall, K. (2012) Guidelines for collecting and maintaining archives for genetic monitoring. *Conservation Genetics Resources,* **4,** 527-536.

Jerde, C.L., Chadderton, W.L., Mahon, A.R., Renshaw, M.A., Corush, J., Budny, M.L., Mysorekar, S. & Lodge, D.M. (2013) Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. *Canadian Journal of Fisheries and Aquatic Sciences,* **70,** 522-526.

Jerde, C.L., Mahon, A.R., Chadderton, W.L. & Lodge, D.M. (2011) "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters,* **4,** 150-157.

Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P.M., Woodcock, P., Edwards, F.A., Larsen, T.H., Hsu, W.W., Benedick, S., Hamer, K.C., Wilcove, D.S., Bruce, C., Wang, X., Levi, T., Lott, M., Emerson, B.C. & Yu, D.W. (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecol Lett,* **16,** 1245-1257.

King, J.R. & Porter, S.D. (2004) Recommendations on the use of alcohols for preservation of ant specimens (Hymenoptera, Formicidae). *Insectes Sociaux,* **51,** 197-202.

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

Krueger, C.C. & May, B. (1991) Ecological and genetic effects of salmonid introductions in North America. *Canadian Journal of Fisheries and Aquatic Sciences,* **48,** 66-77.

Kunin, V., Engelbrektson, A., Ochman, H. & Hugenholtz, P. (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology,* **12,** 118-123.

Lodge, D.M., Williams, S., MacIsaac, H.J., Hayes, K.R., Leung, B., Reichard, S., Mack, R.N., Moyle, P.B., Smith, M., Andow, D.A., Carlton, J.T. & McMichael, A. (2006) Biological Invasions: Recommendations for U.S. Policy and Management. *Ecological Applications,* **16,** 2035-2054.

Lovell, S.J. & Stone, S.F. (2005) The Economic Impacts of Aquatic Invasive Species:A Review of the Literature. *NCEE Working Paper Series*. USEPA.

MacIsaac, H.J. (1996) Potential abiotic and biotic impacts of zebra mussels on the inland waters of North America. *American Zoologist,* **36,** 287-299.

Mahon, A.R., Jerde, C.L., Galaska, M., Bergner, J.L., Chadderton, L., Lodge, D., Hunter, M.E. & Nico, L.G. (2013) Validation of eDNA surveillance sensitivity for detection of Asian carps in controlled and field experiments. *PLoS One,* **8,** e58316.

Martinson, J. & Struewing, I. (2010) Homogenization /Slurry Generation of larval Fish Tissue

Matarese, A.C., Spies, I.B., Busby, M.S. & Orr, J.W. (2011) Early larvae of Zesticelus profundorum (family Cottidae) identified using DNA barcoding. *Ichthyological research,* **58,** 170-174.

Meyer, A. (1993) Evolution of mitochondrial DNA in fishes. *Biochemistry and molecular biology of fishes* (ed. H.a. Mommsen). Elsevier Science Publishers B.V.

Meyer, C.P. & Paulay, G. (2005) DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS Biol,* **3,** e422.

Mills, E.L., J. H. Leach, J. T. Carlton and C. L. Secor (1994) Exotic Species and the Integrity of the Great Lakes. *BioScience,* **44,** 666-676.

Mills, E.L., Leach, J.H., Carlton, J.T. & Secor, C.L. (1993) Exotic Species in the Great Lakes: A History of Biotic Crises and Anthropogenic Introductions. *Journal of Great Lakes Research,* **19,** 1-54.

MNDNR (2014) Invasive Species. (ed. D.o.N. Resources). Minnesota.

Nagy, Z.T. (2010) A hands-on overview of tissue preservation methods for molecular genetic analyses. *Organisms Diversity & Evolution,* **10,** 91-105.

Pimentel, D., Zuniga, R. & Morrison, D. (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics,* **52,** 273-288.

Pochon, X., Bott, N.J., Smith, K.F. & Wood, S.A. (2013) Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. *PLoS One,* **8,** e73935.

Polz, M.F. & Cavanaugh, C.M. (1998) Bias in Template-to-Ratios in Multitemplate PCR. *Appl Environ Microbiol,* **64,** 3724.

Pothoven, S.A., Grigorovich, I.A., Fahnenstiel, G.L. & Balcer, M.D. (2007) Introduction of the Ponto-Caspian bloody-red mysid Hemimysis anomala into the Lake Michigan basin. *Journal of Great Lakes Research,* **33,** 285-292.

Prendini, L., Hanner, R. & DeSalle, R. (2002) Obtaining, Storing and Archiving Specimens and Tissue Samples for Use in Molecular Studies. *Techniques in molecular evolution and systematics* (eds R. Desalle, G. Giribet & W.C. Heeler), pp. 176-248. Birkhaeuser Verlag AG, Basel.

Ratnasingham, S. & Hebert, P.D. (2007) BARCODING BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes.*

Ricciardi, A. (2001) Facilitative interactions among aquatic invaders: is an "invasional meltdown" occurring in the Great Lakes? *Canadian Journal of Fisheries and Aquatic Sciences,* **58,** 2513-2525.

Ricciardi, A. (2006) Patterns of invasion in the Laurentian Great Lakes in relation to changes in vector activity. *Diversity and Distributions,* **12,** 425-433.

Rothlisberger, J.D., Chadderton, W.L., McNulty, J. & Lodge, D.M. (2010) Aquatic invasive species transport via trailered boats: what is being moved, who is moving it, and what can be done. *Fisheries,* **35,** 121-132.

Saccone, C., De Giorgi, C., Gissi, C., Pesole, G. & Reyes, A. (1999) Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene,* **238,** 195-209.

Saunders, G.W. (2009) Routine DNA barcoding of Canadian Gracilariales (Rhodophyta) reveals the invasive species Gracilaria vermiculophylla in British Columbia. *Mol Ecol Resour,* **9 Suppl s1,** 140-150.

Schneider, C., Owens, R., Bergstedt, R. & O'Gorman, R. (1996) Predation by sea lamprey (Petromyzon marinus) on lake trout (Salvelinus namaycush) in southern Lake Ontario, 1982-1992. *Canadian Journal of Fisheries and Aquatic Sciences,* **53,** 1921-1932.

Simon, T.P. & Vondruska, J.T. (1991) Larval identification of the ruffe, Gymnocephalus cernuus (Linnaeus)(Percidae: Percini), in the St. Louis River Estuary, Lake Superior drainage basin, Minnesota. *Canadian journal of zoology,* **69,** 436-442.

Smith, S.H. (1970) Species interactions of the alewife in the Great Lakes. *Transactions of the American Fisheries Society,* **99,** 754-765.

Spies, I., Gaichas, S., Stevenson, D., Orr, J. & Canino, M. (2006) DNA-based identification of Alaska skates (Amblyraja, Bathyraja and Raja: Rajidae) using cytochrome c oxidase subunit I (coI) variation. *J Fish Biol,* **69,** 283-292.

Stein, E.D., Martinez, M.C., Stiles, S., Miller, P.E. & Zakharov, E.V. (2014) Is DNA barcoding actually cheaper and faster than traditional morphological methods: results from a survey of freshwater bioassessment efforts in the United States? *PLoS One,* **9,** e95525.

Stein, E.D., White, B.P., Mazor, R.D., Miller, P.E. & Pilgrim, E.M. (2013) Evaluating Ethanol-based Sample Preservation to Facilitate Use of DNA Barcoding in

Routine Freshwater Biomonitoring Programs Using Benthic Macroinvertebrates. *PLoS One,* **8**.

Stribling, J.B., Pavlik, K.L., Holdsworth, S.M. & Leppo, E.W. (2008) Data quality, performance, and uncertainty in taxonomic identification for biological assessments. *Journal of the North American Benthological Society,* **27,** 906-919.

Trebitz, A., Kelly, J., Hoffman, J., Peterson, G. & West, C. (2009) Exploiting habitat and gear patterns for efficient detection of rare and non-native benthos and fish in Great Lakes coastal ecosystems. *Aquatic Invasions,* **4,** 651-667.

Trebitz, A.S., West, C.W., Hoffman, J.C., Kelly, J.R., Peterson, G.S. & Grigorovich, I.A. (2010) Status of non-indigenous benthic invertebrates in the Duluth–Superior Harbor and the role of sampling methods in their detection. *Journal of Great Lakes Research,* **36,** 747-756.

Urschel, M.R. & O'Brien, K.M. (2008) High mitochondrial densities in the hearts of Antarctic icefishes are maintained by an increase in mitochondrial size rather than mitochondrial biogenesis. *Journal of Experimental Biology,* **211,** 2638-2646.

USGS (2012) Nonindigenous Aquatic Species Database. *http://nas.er.usgs.gov* Gainesville, FL.

Vander Zanden, M.J., Hansen, G.J.A., Higgins, S.N. & Kornis, M.S. (2010) A pound of prevention, plus a pound of cure: Early detection and eradication of invasive species in the Laurentian Great Lakes. *Journal of Great Lakes Research,* **36,** 199-205.

Wagner, A., Blackstone, N., Cartwright, P., Dick, M., Misof, B., Snow, P., Wagner, G.P., Bartels, J., Murtha, M. & Pendleton, J. (1994) Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. *Systematic Biology***,** 250-261.

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

Zhan, A., Hulák, M., Sylvester, F., Huang, X., Adebayo, A.A., Abbott, C.L., Adamowicz, S.J., Heath, D.D., Cristescu, M.E., MacIsaac, H.J. & Pond, S.K. (2013) High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution,* **4,** 558-565.

Zhan, A., Xiong, W., He, S. & Macisaac, H.J. (2014) Influence of artifact removal on rare species recovery in natural complex communities using high-throughput sequencing. *PLoS One,* **9,** e96928.

Zhu, B., Fitzgerald, D., Mayer, C., Rudstam, L. & Mills, E. (2006) Alteration of ecosystem function by zebra mussels in Oneida Lake: impacts on submerged macrophytes. *Ecosystems,* **9,** 1017-1028.

APPENDICES

Appendix A Supplemental information

Appendix B: Additional Experiments and Lessons Learned

Table A.1 Taxonomic nomenclature for species used to construct samples.

		Sample composition		
Experimental sample set	Sample sub-sets	Family	Scientific name	Common name
S1				
	N/A	Gadiidae	*L. lota*	Burbot
		Catostomidae	*Catostomus spp.*	White and Longnose Sucker
		Centrarchidae	*P. nigromaculatus*	Black Crappie
		Percidae	*P. flavescens*	Yellow Perch
		Osmeridae	*O. mordax*	Rainbow Smelt
S2				
	N/A	Centrarchidae	*A. rupestris*	Rockbass
		Gasterosteiidae	*G. aculeatus*	3-Spine Stickleback
		Esocidae	*E. lucius*	Northern Pike
		Cyprinidae	*N. hudsonius*	Spottail Shiner
		Gobiidae	*P. semulinaris*	Tubenose Goby
S3T1 (S3-a,b,c) S3T2 (S3-d)				
	S3-a, b, c, d	Catostomidae	*Catostomus spp.*	White and Longnose Sucker
		Percopsidae	*P. omiscomaycus*	Troutperch
	S3-b, c, d	Gobiidae	*P. semilunarus*	Tubenose Goby
		Percidae	*E. nigrum*	Johnny Darter
		Percidae	*P. flavescens*	Yellow Perch
	S3-c	Esocidae	*E. lucius*	Northern Pike
		Cyprinidae	*N. crysoleucas*	Golden Shiner
		Osmeridae	*O. mordax*	Rainbow Smelt
		Percidae	*P. caprodes*	Logperch
		Centrarchidae	*A. rupestris*	Rockbass
		Centrarchidae	*M. salmoides*	Largemouth Bass

Table A.2 List of reagents used for the **a.** initial PCR amplification of CO1 markers, and **b.** secondary PCR with fusion primers to prepare CO1 amplicons for 454 pyrosequencing.

a. CO1 barcode amplification	**b.** Library adapter preparation with fusion primers
2.0 uL multi-template DNA normalized to 10 ng/uL	5.0 uL purified amplicon DNA normalized to 10 ng/uL
9.9 uL molecular biology grade water	37 uL molecular biology grade water
2.0 uL 10X PCR buffer (Qiagen)	5.0 uL 10X Fast Start Buffer # 2 (454 Life Sciences)
0.6 uL 25mM $MgCl_2$	1 uL 10mM dNTPs
0.4 uL 10X dNTPs	1 uL of each 1.25 mM fusion primer
0.5 uL of forward and reverse primer(s) (Table 4)	1 uL Fast Start High Fidelity Taq (454 Life Sciences)
0.1 uL 10mM taq polymerase (Qiagen)	

Table A.3 Description of primers used to amplify CO1 marker. Highlighted portions of the primer cocktail sequences are M13 tails and were necessary to effectively complete secondary PCR with fusion primers.

Sample set	Primer name	Primer ratio	Cocktail name/Primer sequence 5'-3'	References
S1	dgLCO1490	1	GGTCAACAAATCATAAAGAYATYGG	Folmer *et al.* 1994
	dgHCO2198	1	TAAACTTCAGGGTGACCAAARAAYCA	Folmer *et al.* 1994
S2, S3T1, S3T2	**CO1-3**		**C_FishF1t1-C_FishR1t1**	**Ivanova *et al.* 2007**
	VF2_tl	1	GTAAAACGACGGCCAGT CAACCAACCACAAAGACATTGGCAC	*Ward *et al.* 2005
	FR1d_tl	1	CAGGAAACAGCTATGAC ACCTCAGGGTGTCCGAARAAYCARAA	*Ward *et al.* 2005
	FishF2_tl	1	GTAAAACGACGGCCAGT CGACTAATCATAAAGATATCGGCAC	*Ward *et al.* 2005
	FishR2_tl	1	CAGGAAACAGCTATGAC ACTTCAGGGTGACCGAAGAATCAGAA	Ivanova *et al.* 2007
	M13 fusion tails			
	M13F-21	n/a	GTAAAACGACGGCCAGT	Messing 1983
	M13R-27	n/a	CAGGAAACAGCTATGAC	Messing 1983

*Indicates original reference for the un-tailed version of each primer.

Table A.4 Data pre-processing with Qiime software for Set 1 (S1), Set 2 (S2) and Set 3 Treatment 1 and 2 (S3T1 and S3T2) including parameter description and settings; differences between sample sets reflect modifications aimed at optimizing pre-processing for the purpose of our study.

Quality filter parameters	S1	S2, S3T1, S3T2
Acceptable sequence length	80 - 1000	200 - 1000
Maximum allowed ambiguous bases	6	6
Minimum acceptable mean quality score	22	20
Maximum allowed homopolymer run in base pairs	10	10
Maximum allowed primer mismatches	3	7
Size of quality score window	50	100
Minimum allowed sequence length	80	200
(-z truncate option was enabled so all reads were written with identifiable barcodes by without an identifiable reverse primer	yes	yes
Uncorrected barcodes were not written	yes	yes
Corrected barcodes were written with the appropriate barcode category	yes	yes
Corrected unassigned reads were not written	yes	yes
Total reads associated with valid barcodes that were not in mapping file were written	yes	yes

Table A.5 Qiime BLAST_fragment parameter settings used to identify PCR artifacts (chimeric sequences) for each experimental sample set.

Parameter description	Parameter setting	Description
No. fragments	3	Each sequence is divided in fragments comprised of an equal number of bases
Taxonomic depth	species	Each fragment is blasted against a reference database to the specified depth
Percent similarity	90	Each fragment must be at least this similar to a reference sequence otherwise the fragment is labeled "no blast hit" default setting was used.

Table A.6 UCLUST parameter settings for OTU picking and clustering for all sample Sets. *Percent similarity was 0.98 for Set 1.

UCLUST[1] parameter options	**Parameter settings**	**Parameter setting description**
Application	UCLUST	Program name
Similarity	0.97*	Identity threshold (t) (i.e., >98% similarity = sequences not clustered under the same representative OTU)
Enable reverse strand matching	TRUE	(+) and (-) strand matching (default is (+) strand only)
Exact match	FALSE	Max accepts =1 and max rejects = 0; guarantees that a match will be found if one exists, but not that the best match will be found. Will find match will be found if one exists, but not that the best match will be found.
Maximum accepts	20	Default 20. Keep searching until n hits have been found, then report the best. Default 1. Zero means infinity, i.e. don't stop however many matches have been found (but will still stop if the maximum number of rejects has occurred). Use –maxaccepts 0 –maxrejects 0 to force a search of the entire database with every query, this guarantees that the best hit will be found, if one exists.
Maximum rejects	500	Default 500. Keep searching until n rejects have occurred, then report a failure to find a hit. Zero means infinity, i.e. keep searching until all a hit is found or database sequences have been tested.
New cluster identifier	none	Identifies a new cluster
Optimal match	TRUE	Guarantees that every seed will be aligned to the query, and that every sequence will therefore be assigned to the highest-identity seed that passes the identity threshold (t). All pairs of seeds are guaranteed to have identity < t.

Prefilter identical sequences	TRUE	Sequences that are identical sub-sequences (prefixes) of longer sequences are not considered during the actual clustering.
Presort by abundance	TRUE	Initiates presort by abundance. Companion to 'suppress sort' command. The most abundant sequence is likely to be a true biological sequence, while less common sequences may be artifacts due to sequencing error or PCR artifacts such as chimeras. Since input order is important to picking seeds, presorting by abundance increases the likelihood that seeds represent "true" biological sequences.
Stable sort	TRUE	Specifies that a stable algorithm should be used for U-sorting.
Stepwords	20	Default 20. Step words value to UCLUST. Stepping speeds up database searching. This is effective when the number of words in common between the query and target is expected to be large. Then it is expensive to check all words, and stepping selects a subset of words in the query. This means that the number of query words is chosen so that approximately 20 words are expected to be found in the target sequence. Stepping may reduce sensitivity and may reduce the probability that the best hit is found first.
Suppress sort	TRUE	Suppresses standard presorting method to allow presort by abundance to occur.
Word length	12	Default 12. Word length for unique word index.

1. Edgar, Robert C., Usearch User Guide, version 5.2, 2011, October 15 (http://www.drive5.com/usearch/UsearchUserGuide5.2.pdf)

Appendix B: Additional Experiments and Lessons Learned

B.1 70x75 PicoTiter ™ Plating Format

Plating overview

A 70x75 PicoTiter™ plate can be divided by a gasket into 2, 4, 8 or 16 regions with *n* samples pooled in each region depending on the number of multiplex identifiers (MIDs) used to tag individual samples. Generally, there is an indirect relationship between the number of regions, MIDs per region and the estimated number of sequences (reads) per region (Table B.1.1). The plate layout should be cohesive with study objectives to capitalize on a sequencing run (454LifeSciences 2009).

Plating layout for detection limit evaluation

For each sample Set, PTP layout was designed to minimize non-detection risk at each expected target detection level. Result based modifications were made between Sets as necessary to optimize layout design. Set 1 (S1) samples were plated so the expected target detection level for each sample was less than the manufacturers minimum estimation (e.g., a single sample with an expected target detection level of 1:1000 was plated in a region with an estimated 2,500 – 4,000 reads would be produced for that region). We expected this strategy to generate enough sequences to maximize target detection at low levels. On average, S1 454 run generated 24% fewer reads per sample than the minimum estimation which led to Set 2 (S2) PTP layout revisions. The minimum estimation was reduced by 24% (Table B.1.2) to correct the observed S1 underestimation and samples were plated where the tested detection level was at least 10X less than the lower estimation (e.g., a sample testing a detection level of 1:1000 would be placed where a minimum of 10,000 reads were expected); upper estimations were disregarded. For S2 samples, an average of 29% more sequences than the adjusted lower estimation were generated; for Set 3 Treatment 1 (S3T1) and Treatment 2 (S3T2) PTP layout sequence estimations were adjusted back to the manufacturer's estimates initially used for S1. Similar to S2, S3 samples were plated where the tested detection level was at least 10X less than the lower estimation.

Observed differences between S1 and S2 454 runs could be attributed to the tissue preservation methods used for S1 specimens. Fish tissue used to construct S1 samples was not collected, nor stored in conditions ideal for DNA analysis. Adjustments to the layout strategy resulted in improved run quality for both S2 and S3T1, T2 samples. The PTP layout is an important part of the study design and must be thoroughly considered to minimize non-detection risk.

Table B.1.1 454 GS FLX & FLX+ minimum (s_{min}) and maximum (s_{max}) estimated sequences generated per region for a given gasket format divided into 1, 2, 4, 8, or 16 regions (*r*) with (*n*) total samples per region. For our study a maximum of 10 MID tags were available for multiplexing so *n* ≤ 10. Estimations were used to determine the best way to plate our samples in order to minimize non-detection risk of each target taxon. Set 1 and Set 3, Treatments 1 and 2 PTP layouts were based on values in this table.

r		1		2		4		8		16	
		s_{min}	s_{max}	s_{min}	s_{max}	s_{min}	s_{max}	s_{min}	s_{max}	s_{min}	s_{max}
n	10	30,000	80,000	20,000	40,000	10,000	16,000	5,000	8,000	2,500	4,000
	9	33,333	88,889	22,222	44,444	11,111	17,778	5,556	8,889	2,778	4,444
	8	37,500	100,000	25,000	50,000	12,500	20,000	6,250	10,000	3,125	5,000
	7	42,857	114,286	28,571	57,143	14,286	22,857	7,143	11,429	3,571	5,714
	6	50,000	133,333	33,333	66,667	16,667	26,667	8,333	13,333	4,167	6,667
	5	60,000	160,000	40,000	80,000	20,000	32,000	10,000	16,000	5,000	8,000
	4	75,000	200,000	50,000	100,000	25,000	40,000	12,500	20,000	6,250	10,000
	3	100,000	266,667	66,667	133,333	33,333	53,333	16,667	26,667	8,333	13,333
	2	150,000	400,000	100,000	200,000	50,000	80,000	25,000	40,000	12,500	20,000
	1	300,000	800,000	200,000	400,000	100,000	160,000	50,000	80,000	25,000	40,000

Table B.1.2 Set 2 estimated sequences per sample for a given for each gasket format divided into 1, 2, 4, 8 and 16 regions (*r*) with (n) total samples per region. A maximum of 10 MID tags are available for multiplexing so *n* ≤ 10.

r		1	2	4	8	16
n	10	22,800	15,200	7,600	3,800	1,900
	9	25,333	16,889	8,444	4,222	2,111
	8	28,500	19,000	9,500	4,750	2,375
	7	32,571	21,714	10,857	5,429	2,714
	6	38,000	25,333	12,667	6,333	3,167
	5	45,600	30,400	15,200	7,600	3,800
	4	57,000	38,000	19,000	9,500	4,750
	3	76,000	50,667	25,333	12,667	6,333
	2	114,000	76,000	38,000	19,000	9,500
	1	228,000	152,000	76,000	38,000	19,000

B.2 Investigation into low quality/poor quantity DNA extracted from Set 3 Treatment 2 (S3T2) samples assessing the effect of detritus presence on detection success

Overview

S3 was originally designed to evaluate the effect of extraneous factors contributing to sample complexity. Two factors common to field samples were investigated: Treatment 1 (T1) varying species richness, Treatment 2 (T2) sample residue/detritus presence (Table B.2.1). Samples were constructed using homogenates prepared from larval fish, detritus and TE buffer.

DNA was extracted from T2 samples using the DNeasy ® Blood and Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA extraction success was verified by electrophoresis with 1% agarose gels. Gel results indicated S3T2 DNA template was of sufficient quantity but low quality (Fig. B.2.1). DNA fragments were not of sufficient length for PCR amplification of CO1 marker.

Several hypotheses were formed to help determine the most likely explanation for S3T2 gel results

1) *Something in the detritus interfered with DNA extraction*

2) *Something in the detritus is degrading the DNA*

3) *The DNA was degraded during the creation of the samples*

The first hypothesis seemed unlikely as large amounts of short length DNA was extracted. The second hypothesis also seemed improbable for all samples except control replicates comprised of detritus, because the controls produced low quality and quantity DNA. The third hypothesis also seemed unlikely because S3T1 samples produced DNA template sufficient for PCR and S3T2 samples were constructed using the same methods. Nonetheless, this explanation was more probable and became the premise for additional experiments designed to test the effectiveness of homogenate preservatives, TE buffer and 95% EtOH and DNA kit used for extractions (Table B.2.2).

Gel results (Fig. B.2.2) suggested replicates constructed with 95% EtOH produced better quantity/quality DNA template with both extraction kits. Subsequently S3T2 design was modified to include homogenates prepared with 95% EtOH and DNA was extracted using DNeasy ® Blood and Tissue Kit (Qiagen, Hilden, Germany).

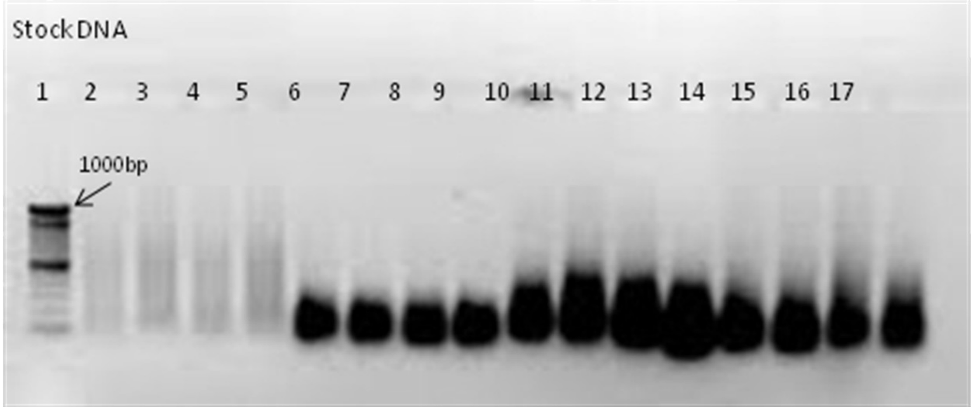


Figure B.2.1 Set 3 Treatment 2 DNA extraction gel results with stock DNA on a 1% agarose gel. Well 1 contains 100 bp DNA ladder; detritus only control in wells 2 – 5; experimental mixes with proportions of detritus to fish biomass equal to 1:1, wells 5 – 9; 1:2, wells 10 – 13; 2:1, wells 14 – 17. Gel results indicate a large quantity of DNA < 1000 bp DNA was extracted from S3T2

samples with larval fish biomass. DNA template was not sufficient for PCR amplification of 650 bp CO1 marker.

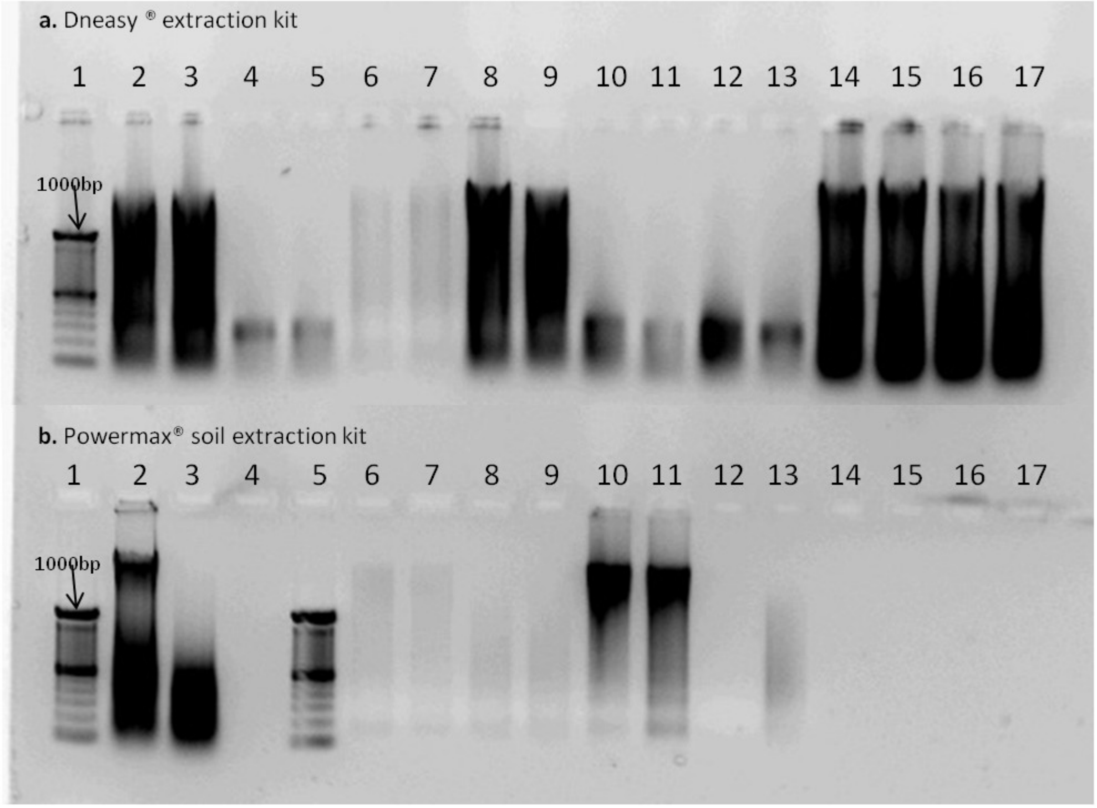


Figure B.2.2 Gel results from Set 3 detritus Treatment comparison assay. **a.** Results from Dneasy ® extraction kit. Well contents: 1) DNA ladder; 2,3) larval fish control, EtOH; 4,5) larval fish control, Tris EDTA; 6,7) detritus control, EtOH; 8,9) 1:1, EtOH; 10, 11) 1:10, Tris EDTA; 12, 13)1:20, Tris EDTA; 14, 15) 1:10, EtOH; 16, 17) 1:20, EtOH. Wells 2, 3,8,9,14 – 17 produced high quality/quantity DNA. **b.** Results from Powermax® soil extration kit. Well contents: 1) DNA ladder; 2,3) larval fish control, EtOH; 4) blank; 5) DNA ladder; 6,7)detritus control, EtOH; 8,9) detritus control, Tris EDTA; 10, 11) 1:1, EtOH; 12, 13) 1:1, Tris EDTA; 14 – 17) blank. Wells 2, 3, 10, 11 produced high quality/quantity DNA.