FINAL REPORT

Development of an Environmental Metagenetics Approach for Monitoring Aquatic Biodiversity

SERDP Project RC-2240

JULY 2017

Michael E. Pfrender
University of Notre Dame

Distribution Statement A
This document has been cleared for public release

SERDP
DOD ▪ EPA ▪ DOE

Page Intentionally Left Blank

This report was prepared under contract to the Department of Defense Strategic Environmental Research and Development Program (SERDP). The publication of this report does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official policy or position of the Department of Defense. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the Department of Defense.

Page Intentionally Left Blank

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
07-07-2017	SERDP Final Report	09/25/2012 – 09/25/17

4. TITLE AND SUBTITLE

Development of an Environmental Metagenetics Approach for Monitoring Aquatic Biodiversity

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Deiner, Kristy: Jerde, Christopher; Evans, Nathan; Lamberti, Gary; Li, Yiyuan; Olds, Brett; Renshaw, Mark; Lodge, David; Pfrender, Michael E.

5d. PROJECT NUMBER
12 RC01-012 / RC02240

5e. TASK NUMBER
1-5

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Notre Dame
Notre Dame, Indiana, USA 46556

8. PERFORMING ORGANIZATION REPORT NUMBER RC-2240

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

DoD Strategic Environmental
Research and Development
Program (SERDP)
Alexandria, VA 22350-3605

10. SPONSOR/MONITOR'S ACRONYM(S)
SERDP

11. SPONSOR/MONITOR'S REPORT NUMBER(S) RC-2240

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution A: unlimited public release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Our overall hypothesis is that genetic-based approaches are more accurate than existing methods in estimating species richness because they are more sensitive in the detection of rare species, whether invasive, threatened, or endangered. Our project was divided into five tasks. Task 1: Develop aquatic environmental DNA (eDNA) metagenetic assays for fishes and amphibians using ultrasequencing and validated for selected species with quantitative real time PCR (qPCR). Task 2: Test the assays on artificial aquatic mesocosm-scale assemblages of known species richness and diversity. Task 3: Test the assays in natural environments whose species richness has been thoroughly evaluated by traditional methods. Task 4: Apply the assays to natural environments with unknown species richness at different spatial scales, for comparison with traditional methods for estimating species richness. Task 5: Apply and compare alternative models for estimating species richness and biodiversity across a variety of aquatic systems. To accomplish these objectives, we developed and applied a new metagenetic toolbox of PCR primers based on multiple mitochondrial loci for the detection via sequencing of freshwater fish and amphibian species richness. By applying these tools to experimental and natural systems we demonstrated that eDNA methods entail much less sampling effort than traditional methods while providing more sensitive estimates of species presence. Therefore, eDNA yielded higher and likely more accurate estimates of species richness than traditional methods.

15. SUBJECT TERMS

Metagenetics, metagenomics, metabarcoding, ultrasequencing, eDNA, species detection, species richness, aquatic biodiversity

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Michael Pfrender
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
UNCLASS	UNCLASS	UNCLASS	UNCLASS	103	(574)631-0591

Page Intentionally Left Blank

Table of Contents

List of Tables

List of Figures

List of Acronyms

Not needed (all acronyms defined in text)

Keywords

Metagenetics, metagenomics, metabarcoding, ultrasequencing, eDNA, species detection, species richness, aquatic biodiversity

Acknowledgements

Authors of this report include Kristy Deiner, Nathan Evans, Chris Jerde, Gary Lamberti, Yiyuan Li, David Lodge, Brett Olds, Michael E. Pfrender and Mark Renshaw. We are grateful to technical assistance from many others, and assistance and guidance from staff at Ft. Custer and Camp Pendleton.

Abstract

Objectives

Our overall hypothesis is that genetic-based approaches are more accurate than existing methods in estimating species richness because they are more sensitive in the detection of rare species, whether invasive, threatened, or endangered. Our work was divided into five tasks. Task 1: Develop aquatic environmental DNA (eDNA) metagenetic assays for fishes and amphibians using traditional polymerase chain reaction (PCR) and ultrasequencing, validated for selected species with quantitative real time PCR (qPCR). Task 2: Test the assays on artificial aquatic mesocosm-scale assemblages of known species richness and diversity. Task 3: Test the assays in natural environments whose species richness has been thoroughly evaluated by traditional methods. Task 4: Apply the assays to natural environments with unknown species richness at different spatial scales, for comparison with traditional methods for estimating species richness. Task 5: Apply and compare alternative models for estimating species richness and biodiversity across a variety of aquatic systems.

Technical Approach

We developed and applied a new metagenetic toolbox of PCR primers for multiple mitochondrial loci for the detection via sequencing of freshwater fish and amphibian species richness. Details for each task follow.

Task 1: To develop markers we used three in silico approaches of literature searches, primer design software, and visual searches of alignments from publicly available sequences. Using experiments, we tested alternative methods for field (DNA sample collection) and laboratory (extraction, library preparation, controls) protocols to increase capture of eDNA, reduce contamination, and achieve room temperature preservation of eDNA. We gathered fish and amphibian specimens from other investigators, museums, and our own collections to create a reference database. We assessed different bioinformatics pipelines with respect to our goal of identifying rare species.

Task 2: We conducted a replicated mesocosm experiment to test the precision of the metagenetic approach for measuring known fish and amphibian species richness; different treatments contained different relative and absolute abundances of the same eight fish species and one amphibian species.

Task 3: We conducted side-by-side sampling with both eDNA and traditional methods in Juday Creek, an Indiana stream with a fish community that has been sampled for many years, and Lawler Pond at Ft. Custer Training Center, Michigan.

Task 4: We conducted eDNA sampling in Eagle Creek at Ft. Custer, and in streams in multiple watersheds at Camp Pendleton, California.

Task 5: We used the bias-corrected Chao II estimator on eDNA incidence data to estimate fish species richness in our mesocosm study, Juday Creek, and Lawler Pond.

Results

Using eDNA assays in experimental mesocosms we detected 100% of stocked species. In a lake and a stream community, using eDNA we detected 100% of the species that were detected by intensive sampling with traditional means (traps, nets, electrofishing), plus additional species not captured. Using Chao estimators, species richness estimates based on eDNA exceeded those based on traditional sampling. Details for each task follow.

Task 1: We identified six primer pairs that are effective for fishes and amphibians. We identified new protocols for DNA sample collection (e.g., use of cellulose nitrate filers, CI extraction), and laboratory protocols including various controls to identify contamination, and discovered that Longmire's preservative provides room temperature preservation for at least two weeks. We created a reference database of sequences for a total of 47 species, including target species, species used in our experiments, and species known to occur at our field sites. Finally, we developed a new bioinformatics pipeline to analyze metagenetic data.

Task 2: In our mesocosm experiment, our metagenetic methods detected all fish and amphibian species in all treatments and all replicates.

Task 3: In Juday Creek, eDNA detected all 12 species caught with traditional methods. eDNA detected four additional species known from the region but never before detected in Juday Creek; these species may be present in the sampled reaches or upstream. From the Ft. Custer pond, eDNA detected all 10 fish species detected with traditional methods plus as many as 11 additional species (depending on the level of bioinformatics stringency we applied).

Task 4: In Eagle Creek, we detected a total of 23 fish species with eDNA, with species number increasing downstream. In Camp Pendleton, we detected 27 native species and 10 exotic species, all previously known from the base. We additionally detected the federally protected Tidewater Goby and the state listed Arroyo chub and Arroyo toad at locations previously known for their occurrence. We did not detect the federally listed Southern Steelhead.

Task 5: Under plausible bioinformatics stringency, species richness estimated using the Chao II estimator from eDNA exceeded the number of species detected with eDNA (and exceeded by even more the number of species detected with traditional methods).

Benefits

The research projects in tasks 1-5 built upon each other. Collectively, these projects demonstrated that eDNA methods entail much less sampling effort than traditional methods while providing more sensitive estimates of species presence. Therefore, eDNA yielded higher, and likely more accurate, estimates of species richness than traditional methods.

Objective

In response to RCSON 12-01 on *Assessment and Monitoring of Biological Diversity: Method Development*, the objective of our research is to improve fundamental and applied understanding of how modern molecular genetic tools can assist management agencies in monitoring the status and trends in native species richness in regions relevant to DoD. We estimate species richness in freshwater habitats and/or for freshwater fish and amphibian species that are difficult or impossible to sample using traditional surveillance methods. Our overall hypothesis is that genetic-based approaches are more accurate than existing methods in estimating species richness because they are more sensitive in the detection of rare species, whether invasive, threatened, or endangered. These approaches also can entail less sampling effort than traditional methods.

When traditional monitoring tools for detection of aquatic species (e.g., nets, traps, acoustic, or electrofishing) are inefficient, impractical, or lacking, management groups are often left in a state of inaction or without accurate species information. In these cases, rare species are often overlooked, leading to errors in inference about existing species richness and biodiversity for a given body of water. We build on recent work showing that the detection of rare species in aquatic systems can be successfully accomplished with molecular genetic detection tools at significantly higher detection probabilities than traditional methods. We proposed to improve and expand current techniques via five tasks:

Task 1) Develop aquatic environmental DNA (eDNA) metagenetic assays for fishes and amphibians using traditional polymerase chain reaction (PCR) and ultrasequencing, validated for selected species with quantitative real time PCR (qPCR);

Task 2) Test the assays on artificial aquatic mesocosm-scale assemblages of known species richness and diversity;

Task 3) Test the assays in natural environments whose species richness has been thoroughly evaluated by traditional methods;

Task 4) Apply the assays to natural environments with unknown species richness at different spatial scales, for comparison with traditional methods for estimating species richness; and

Task 5) Apply and compare alternative models for estimating species richness and biodiversity across a variety of aquatic systems.

Background

The biodiversity of freshwater ecosystems is highly sensitive to human influences (Dudgeon 2010), and thus the status of and trends in aquatic species richness can serve as 'early indicators' of total biodiversity change for a region. Amphibians and fishes are often particularly threatened by environmental changes, featured in indices of biotic integrity, and are of great interest to the public and management agencies (Sala *et al.* 2005; Xenopoulos *et al.* 2005). Assessments of freshwater biodiversity are used by environmental management agencies at local, state, and federal levels and are commonly included in DoD Integrated Natural Resources Management Plans (INRMPs) for military installations (Benton *et al.* 2008). In 2006, the DoD spent $1.6 billion on environmental restoration and conservation at military sites. Freshwater biodiversity monitoring is one of the best ways to prioritize habitats for active management, manage habitats adaptively, and assess the effectiveness of those management efforts (Benton *et al.* 2008).

Aquatic biomonitoring programs traditionally rely on capture or observation of organisms. This approach is often difficult and/or costly due to organism motility and the logistics of underwater sampling. In addition, the risk of harm to species under protection is sometimes high. When tools for species detection are inefficient or lacking, management groups are often left in a state of inaction or are inefficient in their use of existing resources (Jerde *et al.* 2011; Lodge *et al.* 2006; Thompson 2013). In traditional fish monitoring programs that employ nets and/or electrofishing gear, only those organisms present at moderate-to-high abundance are reliably detected (Jerde *et al.* 2011; Magnuson *et al.* 1994). Standard methods for monitoring amphibian diversity and abundance often suffer from similarly low detection probabilities for rare species (Schmidt & Pellet 2009). As the species richness of most ecosystems is dominated by rare species (Magurran 2010), this 'detection gap' for rare species can be systematic and cause errors in inference about species richness and biodiversity and therefore potentially serious mistakes in managing the species under the greatest legal protection by federal and state laws.

To bridge this detection gap, it is essential to develop and adopt monitoring tools with greater detection probabilities for rare species. Improvement in detection capabilities over traditional aquatic sampling methods is the primary objective of our proposal. We propose a synergy between metagenetics and noninvasive genetics, two largely independent fields that are both making rapid advances. Metagenetics is the large-scale assessment of species richness conducted by analyzing homologous genes in bulk environmental samples (Creer *et al.* 2011). Noninvasive genetics is the collection of genetic material from macrofauna without contacting or even seeing the organisms (Beja- Pereira *et al.* 2009). ***Our specific aim in this proposal is to test the hypothesis that shed cellular material suspended in a water body (eDNA) contains a comprehensive genetic inventory of all macrofaunal inhabitants.*** We predict that this inventory is accessible to *en mass* molecular identification via ultrasequencing of aquatic environmental DNA (eDNA).

Task 1) Develop aquatic environmental DNA (eDNA) metagenetic assays for fishes and amphibians using traditional polymerase chain reaction (PCR) and ultrasequencing, validated for selected species with quantitative real time PCR (qPCR)

Methods

Marker choice and primer design

Our focus on fish and amphibians narrowed the choice of DNA markers to the mitochondrial genome (mtDNA), which is well established as the optimal genome for species identification of vertebrates from complex, degraded DNA mixtures such as environmental samples (Taberlet *et al.* 2012). Potential metagenetic markers were developed *in silico* using three strategies:

(1) Searching published literature that used mtDNA markers for species identification of fish, amphibians, or vertebrates.

(2) Application of the metagenetic marker identification software ecoPrimers (Riaz *et al.* 2011) to a custom database we created containing vertebrate mtDNA from NCBI GenBank.

(3) Visually searching alignments, we created containing fish and amphibian mtDNA from NCBI GenBank and OGRe (Organellar Genome Retrieval; http://drake.mcmaster.ca/ogre/).

Our application of these combined strategies resulted in a total of 62 candidate markers distributed across 4 mtDNA regions. These candidate markers were tested on eDNA derived from a pond containing seven fish species. We applied multiple stringent filters to define a set of primer pairs for use in subsequent metagenetic next-generation sequencing. Our filters included i) a temperature gradient PCR using eDNA from a low diversity artificial community to identify an optimal annealing temperature for each primer pair and confirm the production of a PCR product of the expected size, ii) testing PCR with tissue-derived DNA from a wider diversity of fish and amphibian species, iii) evaluation of eDNA PCR product on the Illumina MiSeq platform, and iv) bioinformatics analysis of the Illumina MiSeq data to confirm amplification of target vertebrate taxa.

eDNA sample collection and extraction

As eDNA projects currently employ a variety of capture, preservation and extraction protocols (Lodge *et al.* 2012; Pilliod *et al.* 2013), some standardization could help with comparisons of other aspects of the research that may be heavily influenced by environmental conditions, such as the use of various filter membrane types and pore sizes in the capture of targeted eDNA fragments (Turner *et al.* 2014). The room temperature preservation would additionally allow for application in conditions not suited for cold storage of samples. With these considerations in mind, we conducted a set of four experiments to compare (i) preservation with hexadecyltrimethylammonium bromide (CTAB) or Longmire's buffers among freshly filtered water samples, and samples stored for 1 and 2 weeks at -20, 20 and 45 °C, (ii) the application of the Phenol–chloroform–isoamyl alcohol (PCI) protocol for eDNA extraction from cellulose nitrate filters, polyethersulfone filters, polycarbonate track-etch filters and glass microfiber filters (iii) the PCI DNA extraction protocol with two commercial DNA extraction kits currently featured in eDNA research and (iv) different approaches to the PCI DNA extraction protocol (Table 1).

Table 1. Outline of filtration and extraction experiments, with the treatments evaluated in each experiment (Treatment) and the number of samples analyzed per experimental treatment (N).

Experiment	Treatment	N
Filter preservation	CTAB; fresh	5
	CTAB; −20 °C; 1 week	5
	CTAB; −20 °C; 2 weeks	5
	CTAB; 20 °C; 1 week	5
	CTAB; 20 °C; 2 weeks	5
	CTAB; 45 °C; 1 week	5
	CTAB; 45 °C; 2 weeks	5
	Longmire's; fresh	5
	Longmire's; 20 °C; 1 week	5
	Longmire's; 20 °C; 2 weeks	5
	Longmire's; 45 °C; 1 week	5
	Longmire's; 45 °C; 2 weeks	5
Filter membrane type	0.8 μm; cellulose nitrate (CN)	10
	0.8 μm; polyethersulfone (PES)	10
	1 μm; polycarbonate track-etch (PCTE)	10
	1.5 μm; glass microfibre (GMF)	9
PCI kit comparison	0.45 μm; cellulose nitrate (CN); PCI	10
	0.45 μm; cellulose nitrate (CN); Qiagen	10
	1.5 μm; glass microfibre (GMF); PCI	10
	1.5 μm; glass microfibre (GMF); MoBio	10
DNA extraction	PCI start; ethanol precipitation	10
	PCI start; isopropanol precipitation	10
	CI start; ethanol precipitation	10
	CI start; isopropanol precipitation	10

For all four experiments, 250 mL water samples were collected and filtered immediately through a single filter and filters were subjected to one of several extraction methods fully described in Renshaw et al. 2014: 1) Qiagen's DNeasy Blood and Tissue kit, 2) MoBio's Power-Water DNA Isolation kit, 3) PCI DNA extraction or 4) CI DNA extraction. For this experiment, we used quantitative PCR (qPCR) to assess copy number and assumed that the effects observed for a single species translate to a metagenetic analysis. All DNA extractions were assayed with qPCR Taq-Man primers and probe targeting a 100-bp fragment of the bluegill cytochrome b gene

(Takahara *et al.* 2013) in the following 20 µL mixes: 10 µL of Taq-Man Environmental Master Mix 2.0 (Life Technologies), 1.8 µL of each primer (10 µM stock concentration), 0.25 µL of the Taq- Man probe (10 µM stock concentration), 4 µL of eDNA extract and 2.15 µL of sterile water. The cycling parameters were as follows: a single step at 50 °C for 2 min, a single step at 95 °C for 10 min and 55 cycles at 95 °C for 15 seconds followed by 60 °C for 1 min. DNA copy number was quantified from each experimental replicate using a standard that was created using manufactured DNA at a known quantity. A serial dilution of the standard was run on each qPCR plate and provided a regression line from which the unknown copy numbers of the eDNA extracts could be estimated. All qPCR assays were run on a Mastercycler ep realplex real-time PCR system (Eppendorf) and analyzed with the accompanying realplex 2.2 software. Two negative controls were included on each qPCR plate, both containing the 20 µL mix except for additional sterile water in place of eDNA extract.

ANOVA statistical tests were conducted individually for each of the four experiments to test for differences between mean DNA copy numbers. A two-sided t-test was used to test differences in the average amount of DNA recovered from fresh CTAB and Longmire's extractions within the 'filter preservation experiment'. Technical replicates were averaged for the analysis, residuals from the ANOVAs and t-test were checked for normality using normal Q–Q plots, and pairwise comparisons in the ANOVA were performed using Tukey's post hoc test. All statistics and plots were conducted and created in Mathematica 9.0.1.0 (Wolfram Research, Inc., Version 9.0.1.0, Champaign, IL 2013). All tests conformed to the normality assumptions unless otherwise indicated.

Library preparation

In Appendix C we provide a detail step by step protocol for library preparation and here we provide the basic description of that protocol. Illumina sequencing following a two-stage PCR-based approach as outlined in the Illumina 16S Metagenomic Sequencing Library Preparation guidelines (Illumina, Inc., San Diego, CA, USA). Illumina adaptors as overhangs to the 5' end of each forward (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG) and reverse (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG) primer were added to primer sets in Table 1. The first stage PCR was a 50 µl PCR reaction for each locus-specific amplicon, a single reaction per sample per primer set. We used the following recipe: 29.5 µl sterile water, 10 µl 5x HF buffer, 1 µl 10 mM dNTPs, 1.5 µl 50 mM MgCl2, 1.25 µl 10 µM forward primer, 1.25 µl 10 µM reverse primer, 0.5 µl 2U/µl iProof High-Fidelity DNA Polymerase (Bio-Rad, Hercules, California, USA), and 5 µl DNA. Annealing temperatures (AT) for all primers are provided in Table 1. A "step-down" cycling protocol was incorporated to allow for potential mismatches across a range of taxa. Cycling conditions for all primers were (1) 98 ºC for 2 min; (2) 98 ºC for 10 s; (3) AT1 for 20 s; (4) 72 ºC for 30 s; (5) repeat steps 2-4 an additional 9 times; (6) 98 ºC for 10 s; (7) AT2 for 20 s; (8) 7 2ºC for 30 s; (9) repeat steps 6-8 an additional 9 times; (10) 98 ºC for 10 s; (11) AT3 for 20 s; (12) 72ºC for 30 s; (13) repeat steps 10-12 an additional 29 times; (14) 72 ºC for 10 min; (15) hold at 4 ºC. To confirm PCR products, a 2% agarose gel was run, stained with ethidium bromide, and visualized on a UV light platform. Amplified products were manually excised from the gels with single-use razor blades, cleaned with the QIAquick Gel Extraction Kit (Qiagen, Venlo, Netherlands), and eluted from spin columns with 30µl of Buffer EB.

To complete the Illumina sequencing library and individually barcode each sample, a 50 µl PCR reaction was used for a second stage PCR, consisting of 22 µl sterile water, 10 µl 5x HF buffer, 1 µl 10 mM dNTPs, 1.5 µl 50 mM MgCl2, 5 µl 10 µM Nextera Index Primer 1 (N701-N712), 1.25

µl 10 µM Nextera Index Primer 2 (S502-S508 and S517), 0.5 µl 2U/µl iProof High-Fidelity DNA Polymerase (Bio-Rad, Hercules, California, USA), and 5 µl amplified DNA from the first stage. For the 2nd Stage PCR, the template DNA was a pool of 25ng of DNA derived from all markers from each individual sample.

Temperature cycling conditions for the 2nd Stage PCR consisted of an initial denaturation step at 98 °C for 2 min; followed by 8 cycles of denaturation at 98 °C for 10 sec, annealing at 55 °C for 20 sec, and extension at 72 °C for 30 sec; followed by a final extension step at 72 °C for 10 min. The PCR Clean-Up 2 protocol was followed (16S Metagenomic Sequencing Library Preparation) and DNA concentrations were quantified with the Qubit dsDNA HS Assay (Life Technologies, Carlsbad, California, USA). All four amplicon sizes were verified within each library on a Bioanalyzer DNA 7500 chip (Agilent Technologies, Santa Clara, California, USA). Paired-end Illumina MiSeq sequencing was performed in a single MiSeq run by the University of Notre Dame's Genomics and Bioinformatics Core Facility (http://genomics.nd.edu/) with a MiSeq Reagent Kit v 3 (600-cycle; Illumina, San Diego, California, USA).

Contamination and error controls

To observe potential artifacts such as contamination and errors from PCR, sequencing and bioinformatics, several types of controls were used throughout each study, but generally include field negative controls, extraction negative and positive controls, and PCR negative and positive controls (Goldberg *et al.* 2016). Field negative controls consisted of reverse osmosis (RO) water filled in a sampling container and carried into the field alongside that of the collected eDNA samples. It was processed alongside that of field samples for every step and therefore serves as a full process negative control. A mock community sample was constructed and run through the DNA extraction process alongside eDNA samples and therefore was a positive control for the extraction and amplification process, but doubled as s negative control to detect laboratory and bioinformatic errors. The mock community sample was composed of equal amounts of tissue derived DNA (measured with Qubit) from six Indo-Pacific marine fishes: *Amphiprion ocellaris, Salarias fasciatus, Ecsenius bicolor, Centropyge bispinosa, Pseudanthias dispar*, and *Macropharyngodon negrosensis* for all sites accept Camp Pendleton because this site included ocean samples from the Pacific Ocean. We therefore used five freshwater endemic fish species from south eastern United States: *Umbra limi, Thoburnia atripinnis, Erimyzon sucetta, Notropis topeka,*and *Noturus taylori*. Negative controls consisted of RO water in place of any eDNA sample and were instituted at the extraction and PCR stages of laboratory analysis.

Bioinformatics pipeline

Next generation sequencing provides large volume of data. Analyzing large sequencing data sets in a fast and accurate way is essential. Our goal was to build a bioinformatics pipeline to analyze the Illumina MiSeq data produced from amplicon sequencing using our metagenetic primers. The basic steps of the pipeline that we implemented are illustrated in Figure 1.

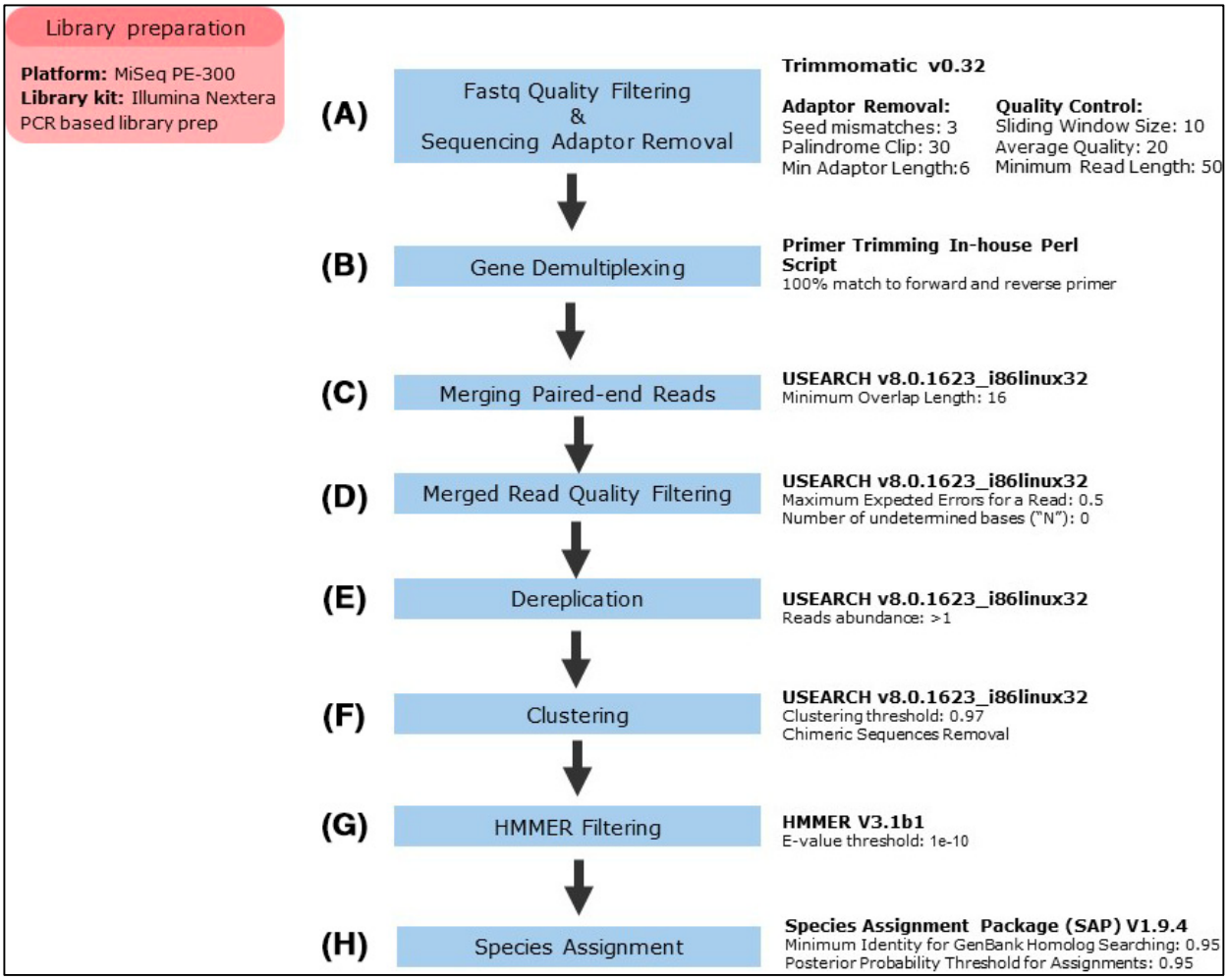


Figure 1: Workflow of the eDNA pipeline designed for analysis with our metagenetic approach.

The pipeline filters raw sequence reads using Trimmomatic v0.32 to remove Illumina sequencing adaptor, low-quality sequences with average quality less than Q20 in any 10-bp window and short sequences with length less than 50-bp (Bolger et al. 2014) (Step A, Figure 1). The pipeline then demultiplexes paired-end reads based on the forward and reverse metagenetic primers (Table 1), while retaining the integrity of each fastq file from read 1 and read 2. We remove sequence reads without an exact match to their expected primer sequences and trimmed primer sequences from all reads (Step B). The pipeline then merges overlapping paired-end reads with USEARCH v8.0.1623 with default settings (Edgar 2010) (Step C). We discard reads with expected errors >0.5 or ambiguous base pairs at any nucleotide site (Step D).

After merging and quality control, the pipeline pools together the merged reads from all samples but retains each marker individually (Step E). Identical sequences are then counted and a file is produced for each gene with all unique merged reads. Merged reads with a count of one are removed. The pipeline then uses USEARCH v8.01623 (Edgar 2010) to cluster sequences within 97 % sequence similarity into Operational Taxonomic Unites (OTUs) and removes potential chimeric sequences generated during PCR (Step F). The OTUs are then summarized by a single centroid sequence to represent the cluster and the count of merged reads clustered into each OTU are counted. The pipeline then removes non-targeted organisms (e.g., bacteria, protists) merged

reads by matching the OTU centroid sequence to a profile hidden Markov model, which uses a position-specific scoring system to estimate the degree of conservation at each site of the reference sequences from our gene regions (Table 1) with HMMER v3.1b (Wheeler & Eddy 2013) (Step G). Lastly, the centroid sequences from OTUs are assigned taxonomic identifications based on the NCBI NR database using SAP v1.9.4 (Munch *et al.* 2008) (Step H).

Reference database generation for targeted taxa

Reference databases are essential for taxonomic assignment of DNA sequences from environmental samples. The utility of a reference database for a specific metagenetic analysis depends on the choice of DNA markers, the taxonomic focus, and the representation of intra- and interspecific genetic diversity. Our goal was to develop a reference database that would represent the target fish and amphibian species considered threatened and endangered at mainland DoD installations (39 fish, 13 amphibians), all species included in our mesocosm experiment (8 fish, 1 amphibian), and all species known to occur in our field sites at Juday Creek and Ft. Custer Training Center. Our goal was to obtain at least one mitochondrial DNA sequence for each of the six DNA markers for each target species by combining publicly available sequences with sequences we generated using available tissues.

A summary of the reference database is given in Appendix A1. The species are broken into two groups, fish and amphibian. For each species, the justification for its inclusion in the database is given ("TE" for threatened or endangered, "ME" for mesocosm, "JC" for Juday Creek, and "FC" for Fort Custer) along with the GenBank accession numbers for each of the six markers. GenBank accessions from outside sources are highlighted in peach, accessions generated in-house are highlighted in gray, and species/marker combinations with no sequence data available are highlighted in red. Tissue samples were collected for 34 (of 39) fish and 13 (of 13) amphibian species from the threatened and endangered list; tissues were collected for 8 (of 8) fish and 1 (of 1) amphibian species from the mesocosm list; tissue samples were collected for 19 (of 23) fish species from the Juday Creek list; and tissues were collected for 16 (of 28) fish and 6 (of 19) amphibian species from the Fort Custer list.

Validation with qPCR

We did not do any head-to-head comparisons of qPCR vs. metagenetics as imagined when we wrote the proposal. The reason for this is because early on it became abundantly clear that we could detect 100 % of species with our metagenetic approach in our controlled experiment (mesocosms), and subsequent field tests with known biodiversity (Ft. Custer and Juday Creek). We therefore focused our efforts in these directions.

Results and discussion

Marker choice and primer design

After the elimination of suboptimal primer pairs at each successive step we retained a robust set of six primer pairs in our metagenetic toolbox (Table 2). All three of our *in silico* development strategies (literature search, ecoPrimers, and visual alignment search) are represented in the final suite of markers. We recommend the primer sets in Table 2 for use in metagenetic studies of fish and amphibians. Use of all six is not necessary and the number of loci and primers sets evaluated for any given monitoring effort should be based on the goals and questions being addressed.

Table 2. Primer sets for PCR amplification of fish/amphibian metagenetic markers from eDNA.

Names (F/R)	target gene	forward primer	reverse primer	amplicon length (bp)	annealing temp (°C) AT_1, AT_2, AT_3	source
L14912/ H15149c	Cyt B	AAAAACCACCGTTGT TATTCAACTA	GCCCCTCAGAATG ATATTTGTCCTCA	413	60°, 58°, 55°	*(Burgener & Hübner 1998)*
Ac12s	12s	ACTGGGATTAGATAC CCCACTATG	GAGAGTGACGGGC GGTGT	385	63°, 60°, 58°	*Current study*
Am12s	12s	AGCCACCGCGGTTAT ACG	CAAGTCCTTTGGG TTTTAAGC	241	65°, 62°, 60°	*Current study*
Ac16s	16s	CCTTTTGCATCATGAT TTAGC	CAGGTGGCTGCTT TTAGGC	330	63°, 60°, 58°	*Current study*
Ve16s	16s	CGAGAAGACCCTATG GAGCTTA	AATCGTTGAACAA ACGAACC	310	65°, 62°, 60°	*Current study*
L2513/H 2714	16s	GCCTGTTTACCAAAA ACATCAC	CTCCATAGGGTCT TCTCGTCTT	202	60°, 58°, 55°	*(Kitano et al. 2007)*

eDNA sample collection and extraction

The main goal of this study was to evaluate eDNA preservation and extraction methods for filtered macrofaunal eDNA, with the potential broad application for studies in a variety of aquatic environments across DOD installations. For the Filter preservation experiment, all replicates amplified and were incorporated into the statistical analyses for all twelve of the experimental treatments (Table 1). For the CTAB preservation buffer, relative to fresh samples, Tukey's post hoc comparisons of the ANOVA results revealed a significantly higher DNA copy number in samples stored at all the three temperatures (-20, 20 and 45 °C) following the 2-week time interval (Figure 2a–c). For the Longmire's preservation buffer, the same result was observed for the 45 °C temperature (Figure 2e), but no significant difference in copy number existed between fresh samples and those stored at 20 °C (Figure 2d). A two-sided t-test of the fresh extractions revealed a significantly higher yield in DNA copy number for the Longmire's preservation buffer as compared to the CTAB preservation buffer (P-value < 0.001; Figure 2f). The Longmire's preservation buffer performed well and we recommend its use going forward because it provides researchers with a room temperature storage buffer that adequately handles elevated temperatures (up to 45 °C tested in this study), and the assimilation of the Longmire's preservation buffer into a PCI DNA extraction protocol has the potential to simultaneously reduce per sample costs and increase the recovery of targeted eDNA fragments thus making this workflow cost effective for managers.

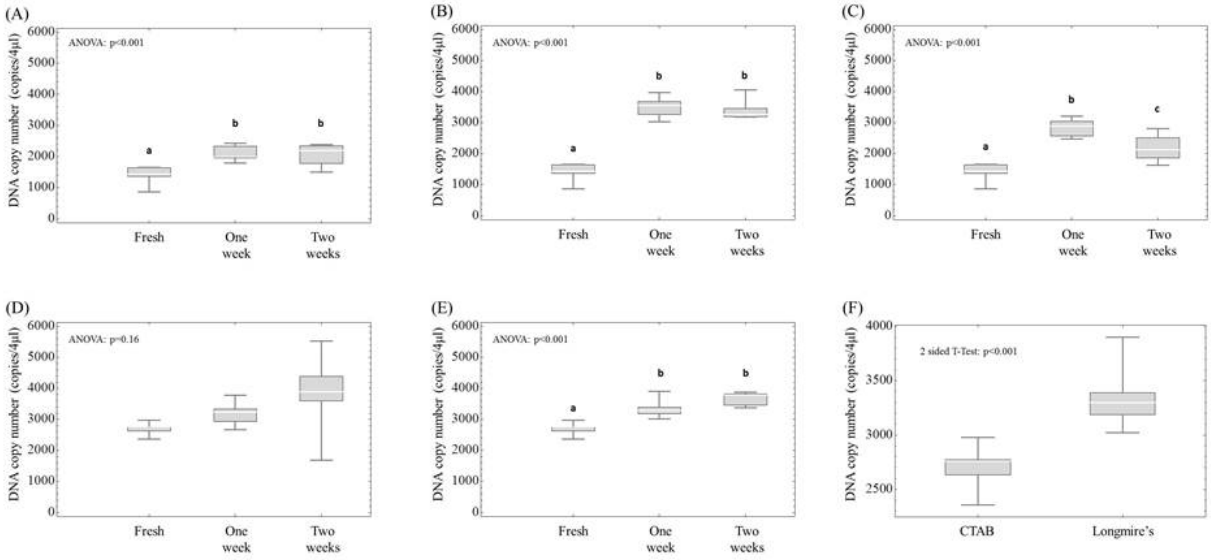


Figure 2. Filter preservation experiment results. Box plot where the top and bottom of the whiskers represent the maximum and mini-mum values, the top and bottom of the boxes represent the 75% and 25% quartiles, and the lines inside the boxes represent the median values. Significance in pairwise comparisons of treatments is noted by letters a, b and c where different letters represent statistically significant differences. Two preservation buffers, CTAB and Longmire's, were evaluated over a 2-week interval of time. (a) CTAB with - 20 °C storage, (b) CTAB with 20 °C storage, (c) CTAB with 45 °C storage, (d) Longmire's with 20 °C storage, (e) Longmire's with 45 °C storage and (f) comparison between CTAB and Longmire's for fresh extractions.

For the PCI kit comparison experiment, all 10 of the samples amplified and were incorporated into the statistical analyses for each of the four experimental treatments (Table 1). Tukey's post hoc comparisons of the ANOVA results revealed that the CN filter with PCI extraction yielded significantly more copies of DNA than the other three experimental treatments; the GMF filter with the MoBio extraction yielded significantly more copies of DNA than both the GMF filter with PCI extraction and the CN filter with Qiagen extraction, which were not significantly different from one another (Figure 3a).

For the DNA extraction experiment, all 10 of the samples amplified and were incorporated into the statistical analyses for each of the four experimental treatments (Table 1). Tukey's post hoc comparisons of the ANOVA results revealed no statistically significant differences among the four experimental treatments (Figure 3b).

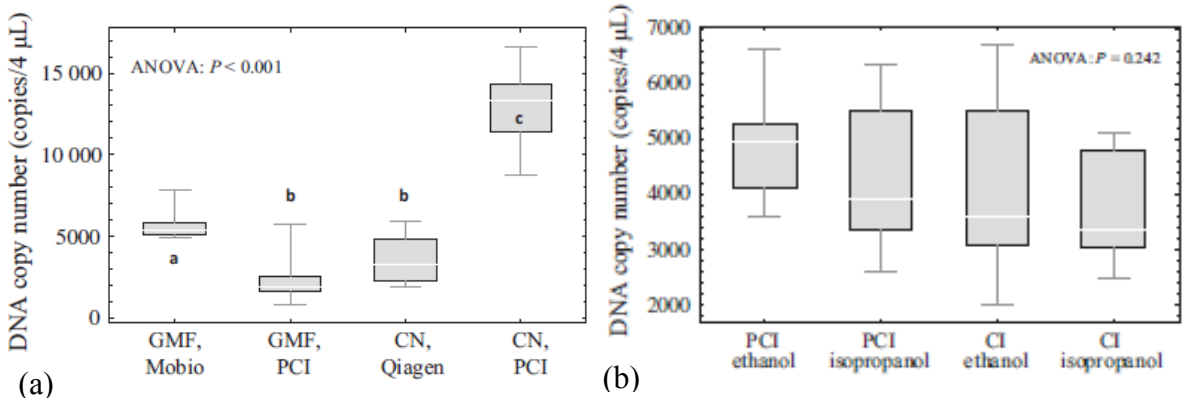


Figure 3. Filter type and extraction method comparisons. (a) Box and whisker plots for the PCI-kit comparison experiment. The top and bottom of the whiskers represent the maxi-mum and minimum values, the top and bottom of the boxes represent the 75% and 25% quartiles, and the lines inside the boxes represent the median values. Significance in pairwise comparisons of treatments is noted by letters a, b and c where different letters represent statistically significant differences between experimental treatments. The four treatments were 1.5-lm glass microfiber filters (GMF) with MoBio extraction, 1.5-lm glass microfiber filters (GMF) with PCI extraction, 0.45-lm cellulose nitrate filters (CN) with Qiagen extraction and 0.45-lm cellulose nitrate filters (CN) with PCI extraction. (b) Box and whisker plots for the DNA extraction experiment. The top and bottom of the whiskers represent the maximum and minimum values, the top and bottom of the boxes represent the 75% and 25% quartiles, and the lines inside the boxes represent the median values. There was no statistical significance in pairwise comparisons between the four experimental treatments: PCI extraction with ethanol precipitation, PCI extraction with isopropanol precipitation, CI extraction with ethanol precipitation and CI extraction with isopropanol precipitation.

Based on these results (Figures 2 & 3), we decided to use a Cellulose Nitrate filter (CN) combined with a CI extraction on our water samples for all sites evaluated in this project. We chose a CI protocol because it reduced the use of a highly toxic chemical and our work did not show that there was a significant effect in the copy numbers detected. However, methods for concentrating and extracting eDNA from environmental samples is a continually developing aspect of the field and we recommend doing a pilot study in new locations as other factors not tested here (PCR inhibitors, pH, etc.) are likely to change detection rates (Goldberg *et al.* 2016)

Library preparation

During the project and while developing the library preparation protocol, methods were switched from using a single index True Seq library preparation method (Evans *et al.* 2016) to a duel index Nextera method (Olds *et al.* 2016) because of problems associated with de-multiplexing pooled samples. While this does not entirely solve the problem, this is an ongoing area of research by the user community of the Illumina high-through-put sequencing platform and until a new solution is apparent, we recommend following the advice of the manufacturer and use the Illumina 16S Metagenomic Sequencing Library Preparation guidelines (Illumina, Inc., San Diego, CA, USA) and that of our protocol provided in Appendix C.

Contamination and error controls

Details pertaining to the utility of controls and how they were used to monitor contamination are described within each study and are outlined in Tasks 3 and 4.

Bioinformatics pipeline

The eDNA pipeline code is available from the website https://github.com/pfrender-laboratory/epps and the pipeline can automatically process multiple sequencing libraries and multiple primers from quality filtering to species assignment (Figure 1).

Reference database generation

Of the 52 threatened and endangered species found on DoD installations in the mainland U.S., we have obtained tissues for 47 species. Thus far, we have been unable to locate a source of tissues for five fish species. We have acquired tissues from mesocosm species, Juday Creek species, and Ft. Custer species. We validated the efficacy of our primer pairs for these species through PCR and Sanger sequencing. A summary of these efforts is shown in Appendix A1. For species labeled with "NGS mitogenome" we sequenced the entire mitochondrial genome as part of an effort to develop a low-cost mitogenome sequencing protocol to enhance reference databases for metagenetic analysis.

Task 2) Test the assays on artificial aquatic mesocosm-scale assemblages of known species richness and diversity

Methods

We conducted a replicated mesocosm experiment to test the precision of the metagenetic approach for measuring fish and amphibian species richness. The research question we investigated was: ***Can eDNA metagenetics detect all fish and amphibian species in mesocosm assemblages with differing species densities and relative abundances?*** The experimental design consisted of four, experimental treatments of fish and amphibian community structure in isolated 206-L tanks: (1) high total density and even relative abundance, (2) low total density and even relative abundance, (3) high total density and skewed relative abundance, (4) low total density and skewed relative abundance. This crossed experimental treatment design enabled us to evaluate the effects of both density (high vs. low) and relative abundance (even vs. skewed) on species detection.

Filtered water was extracted from day 5 of the mesocosm experiment and PCR-amplified using the primer sets in Table 2. PCR conditions, amplicon purification, Illumina library construction, Illumina MiSeq sequencing, and bioinformatics analysis are detailed in task 1. Resulting eDNA sequences were searched for matches to the nine mesocosm study species by mapping them against a reference database of independently-generated amplicon sequences from each study species and primer set.

Results and discussion

In our mesocosm experiment, we detected all species (eight fishes and one amphibian species) in all treatments and all replicates (Table 2). The eDNA metagenetic approach accurately measured the species richness of each community assemblage irrespective of differences in the relative abundance and density of the constituent species.

Table 3. Biomass (g) and number (in parentheses) of each of the nine study species in the experimental mesocosms.

	High Density, Even Abundance			Low Density, Even Abundance			High Density, Skewed Abundance			Low Density, Skewed Abundance		
Species	Tank 1	Tank 2	Tank 3	Tank 1	Tank 2	Tank 3	Tank 1	Tank 2	Tank 3	Tank 1	Tank 2	Tank 3
Campostoma anomalum	29 (10)	36.5 (10)	32.5 (10)	19.7 (4)	17.3 (4)	15.6 (4)	69.1 (18)	72.3 (18)	15.3 (4)	21.1 (5)	19.9 (5)	7.3 (2)
Catostomus commersonii	22.2 (10)	17.4 (10)	20.9 (10)	8.0 (4)	7.1 (4)	10.2 (4)	7 (4)	9.0 (4)	7.3 (4)	3.1 (2)	4.4 (2)	3.5 (2)
Cyprinus carpio	74.8 (10)	88.4 (10)	139.1 (10)	16.3 (4)	19.0 (4)	77.2 (4)	26.4 (5)	26.6 (5)	44.2 (5)	4.4 (2)	36.2 (2)	18 (2)
Fundulus notatus	12.6 (10)	10.8 (10)	14.0 (10)	3.7 (4)	5.0 (4)	5.4 (4)	13.1 (7)	11.1 (7)	5.5 (4)	4.2 (3)	6.7 (3)	3.8 (2)
Gambusia holbrooki	0.7 (10)	0.3 (10)	2.0 (10)	0.8 (4)	0.4 (4)	0.8 (4)	0.5 (4)	0.4 (4)	1.0 (7)	0.1 (2)	0.2 (2)	0.6 (3)
Lepomis macrochirus	10.8 (10)	12.9 (10)	16.0 (10)	4.0 (4)	4.7 (4)	3.7 (4)	5.5 (4)	6.3 (4)	19.9 (18)	1.8 (2)	2.6 (2)	6.4 (5)
Pimephales promelas	6.7 (10)	14.0 (10)	18.4 (10)	3.4 (4)	4.5 (4)	6.5 (4)	67.1 (46)	73.7 (46)	78.6 (46)	28.5 (18)	25.4 (18)	29.9 (18)
Rana catesbeiana	44.5 (10)	47.7 (10)	53.3 (10)	17.2 (4)	18.1 (4)	24.8 (4)	17.4 (4)	21.7 (4)	14.5 (4)	8.8 (2)	8.1 (2)	8.4 (2)
Semotilus atromaculatus	13.4 (10)	39.6 (10)	11.1 (10)	2.7 (4)	13.1 (4)	4.9 (4)	13.6 (4)	15.9 (4)	6.6 (4)	2.9 (2)	4.8 (2)	3.4 (2)

Task 3) Test the assays in natural environments whose species richness has been thoroughly evaluated by traditional methods

Methods

To test the eDNA metagenetic approach in natural communities we conducted side-by-side sampling with both eDNA and traditional methods. The research question we planned to investigate was: ***Can eDNA metagenetics detect at least 70% of the fish and amphibian species in small natural communities?*** We answered this question by conducting studies at two sites as detailed below.

Site 1: Juday Creek, St. Joseph County, Indiana, USA

Environmental DNA samples were first collected from Juday Creek and then traditional sampling was immediately conducted to add a 17th year of survey data (Shirey *et al.* 2016). The reach was first divided into four sections with block nets below and above each section (Figure 4). From each section, two 250 mL water samples were collected one from the bottom of the section and the second from the top of the section four a total of 4 L from the reach. For electrofishing, each section was surveyed with triple-pass backpack electrofishing (Figure 4). Environmental DNA was extracted, sequenced, and bioinformatically analyzed according to the methods described in Task 1 except we only evaluated four of the six primer sets in Table 2 (Ac12s, Ac16s, Am12s and L14735/H15149).

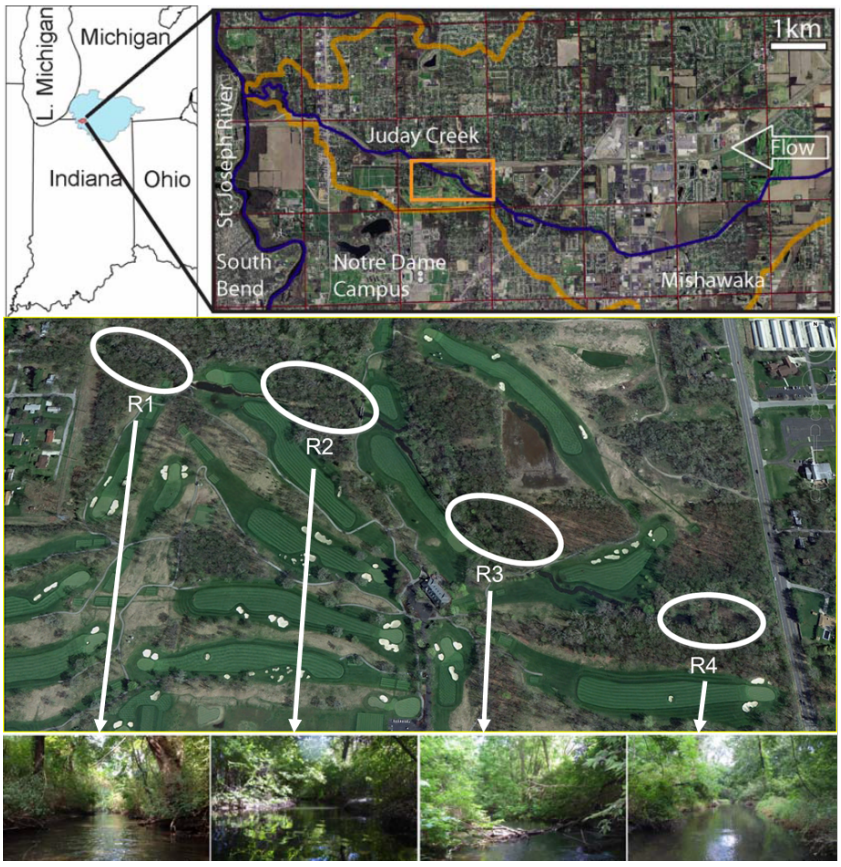


Figure 4. Juday Creek is a groundwater-fed, 3rd-order tributary within the St. Joseph River watershed (highlighted in light blue), which flows into Lake Michigan; the 19-km. stream drains an area of 98 km^2 in northwestern Indiana (41°42'N,86°13'W;elevation = 206m). White ovals indicate sections of the reach that were electrofished across 17 years (including 2013) and sampled for eDNA in 2013. Figure modified from Shirley et al. 2016.

To assign species taxonomic information to reads obtained from sequencing, two different programs were used, SAP v1.9.3 (Statistical Assignment Package; (Munch *et al.* 2008) and

USEARCH v8.0.1623 (Edgar 2010). First, SAP was used to assign OTUs with no a priori knowledge of existing species presence, utilizing all sequences found on the NR database of GenBank. SAP relies on the phylogeny of homologs found in the GenBank database; therefore, species with hybrids in the GenBank database are always assigned to a higher taxonomic level with low posterior probability, such as common carp (*Cyprinus carpio*). Therefore, we combined SAP results with a second method using a global alignment (USEARCH) based on a reference list of species known from Juday Creek along with many related species for which sequence data was available (Table 4).

Table 4. List of species for which we had a reference sequence available for species assignment for each of the four markers used.

Species	**Ac12s**	**Ac16s**	**Am12s**	**L14735/H15149**
Acipenser brevirostrum	X	X	X	X
Acipenser medirostris	X	X	X	X
Acipenser oxyrhinchus	X	X	X	X
Acipenser oxyrhinchus desotoi	X	X	X	X
Acipenser ruthenus	X	X	X	X
Alosa alabamae	X	X	X	X
Alosa sapidissima	X	X	X	X
Ambloplites rupestris	X	X	X	X
Ameiurus melas	X		X	
Ameiurus natalis	X	X		X
Amoya chusanensis		X		X
Aphredoderus sayanus	X	X	X	X
Campostoma anomalum	X	X	X	X
Campostoma ornatum	X		X	
Catostomus commersonii	X	X	X	X
Cottus bairdii	X	X	X	X
Cottus cognatus	X		X	X
Cottus hangiongensis	X	X	X	X
Ctenopharyngodon idella	X	X	X	X
Cyprinella callitaenia	X	X	X	
Cyprinella lutrensis	X	X	X	X
Cyprinella zanema	X	X	X	X
Cyprinodon macularius	X	X	X	X
Cyprinodon rubrofluviatilis	X	X	X	X
Cyprinodon tularosa	X	X	X	X
Cyprinus carpio	X	X	X	X
Elassoma evergladei	X		X	X
Elassoma okatie	X		X	X
Erimyzon oblongus	X	X	X	X
Erimyzon sucetta	X	X	X	X
Esox americanus				

Esox reichertii	X	X	X	X
Etheostoma caeruleum	X	X	X	X
Etheostoma cragini	X	X	X	X
Etheostoma ditrema				X
Etheostoma exile	X	X	X	X
Etheostoma maculatum		X		X
Etheostoma microlepidum		X		X
Etheostoma nigrum	X	X	X	X
Etheostoma okaloosae		X		X
Etheostoma radiosum	X	X	X	X
Etheostoma tuscumbia	X	X	X	X
Eucyclogobius newberryi	X	X	X	X
Fundulus diaphanus	X	X	X	X
Fundulus notatus	X	X	X	X
Gambusia affinis	X	X	X	X
Gambusia holbrooki	X	X	X	X
Gasterosteus aculeatus	X	X	X	X
Gasterosteus aculeatus williamsoni	X	X	X	X
Gila bicolor mohavensis	X	X	X	X
Gila conspersa	X	X	X	X
Gila orcutti	X	X	X	X
Homo sapiens	X	X	X	X
Hypophthalmichthys molitrix	X	X	X	X
Hypophthalmichthys nobilis	X	X	X	X
Ictalurus punctatus	X	X	X	X
Iotichthys phlegethontis	X	X	X	X
Lepomis cyanellus	X	X	X	X
Lepomis gibbosus	X	X	X	X
Lepomis gulosus		X		X
Lepomis macrochirus	X	X	X	X
Lepomis microlophus		X		X
Meda fulgida	X	X	X	X
Microphis brachyurus	X	X	X	X
Micropterus dolomieu	X	X	X	X
Micropterus salmoides	X	X	X	X
Myxocyprinus asiaticus	X	X	X	X
Neosalanx taihuensis	X	X	X	X
Notemigonus crysoleucus	X	X	X	X
Notophthalmus perstriatus				
Notophthalmus viridescens				
Notropis anogenus				X
Notropis heterodon	X	X	X	X

Notropis stramineus	X	X	X	X
Notropis topeka	X	X	X	X
Noturus taylori	X	X	X	X
Oncorhynchus clarkii henshawi	X	X	X	X
Oncorhynchus clarkii stomias	X	X	X	X
Oncorhynchus keta	X	X	X	X
Oncorhynchus kisutch	X	X	X	X
Oncorhynchus mykiss	X	X	X	X
Oncorhynchus tshawytscha	X	X	X	X
Oreochromis aureus	X	X	X	X
Perca flavescens	X	X	X	X
Percina cymatotaenia				X
Percina macrolepida	X	X	X	X
Percina rex	X	X	X	X
Phoxinus eos	X	X	X	X
Phoxinus phoxinus	X		X	
Pimephales notatus	X	X	X	X
Pimephales promelas	X	X	X	X
Pomoxis annularis		X		X
Pomoxis nigromaculatus	X	X	X	X
Rhinichthys atratulus	X	X	X	X
Rhynchocypris kumgangensis	X	X	X	
Rhinichthys obtusus	X	X	X	X
Salmo trutta	X	X	X	X
Salvelinus confluentus		X		
Salvelinus leucomanis		X		
Semotilus atromaculatus	X	X	X	X
Squaliobarbus curriculus				X
Thaleichthyes pacificus	X	X	X	X
Thoburnia atripinnis	X	X	X	X
Tiaroga cobitis	X	X	X	X
Umbra limi	X	X	X	X
Umbra pygmaea	X	X	X	X

A consensus species assignment was made between the two methods by comparing the assignment from SAP with the global identity made by USEARCH and determined if the taxonomic assignment was the same. If not, we conducted a manual verification of the target sequence by comparing it to a reference sequence (Figure 5). OTUs that could not be assigned to species level were excluded from further analysis. A species was considered detected in a sample if we detected two or more reads in two or more markers.

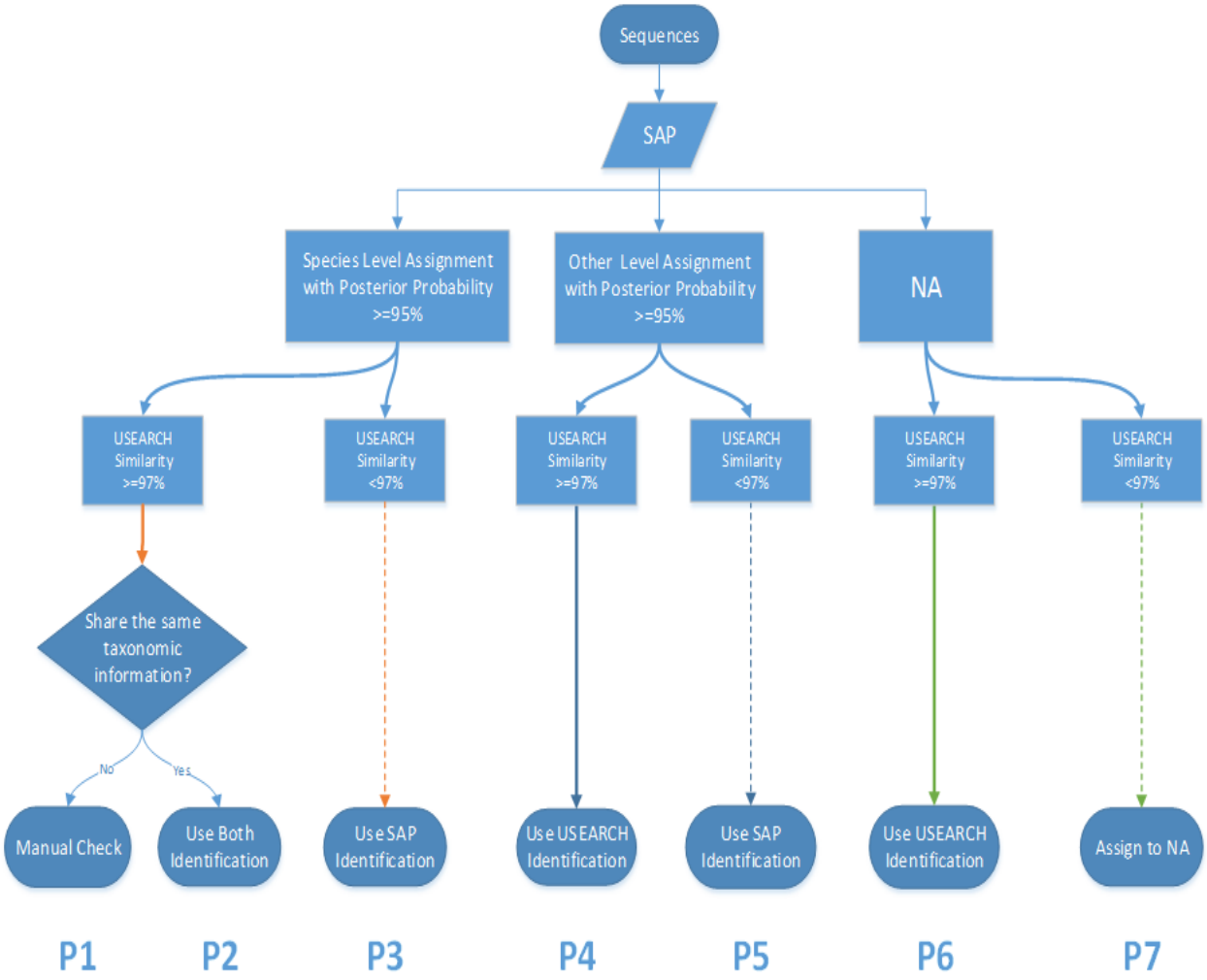


Figure 5. Decision making flowchart of species assignment utilizing OTU sequences input into SAP and USEARCH. For example, in the case of path 1 (P1): if SAP provides a species-level assignment with posterior probability >=95%, USEARCH has a global alignment with identity >=97%, and the species assignments from the two approaches are identical and the species assignment was used. If the two assignments from each program are different (P2), we manually check the assignment against GenBank NR database and manually made the decision as to the appropriate species assignment based on the similarity of the sequence to the reference.

Following species assignment of reads, contamination and its potential influence on our estimated species richness was assessed by fitting a Poisson distribution of contaminant DNA found in concurrently run negative controls and then evaluated the probability that a sequence matching a species in each sample could have arisen by chance. We then flagged any species marker combination with a probability >0.001 as a false positive detection due to contamination (full details see Appendix A2). In cases where a sample had potential contamination for a marker, we re-evaluated our conclusion about positive detection and required reads with the assigned species to be present in at least two markers with at least two reads.

Site 2: Lawler pond, Fort Custer, Michigan Army National Guard, Kalamazoo County, Michigan, USA

Environmental DNA samples were taken from Lawler Pond and then immediately traditional sampling was conducted. On June 1, 2014, one day prior to the start of our traditional sampling, we collected one 250-mL water sample from each of 30 locations distributed throughout Lawler Pond (Figure 6). In addition, we collected one 250-mL water sample from the stream inflow into Lawler Pond (Figure 6). Each water sample was collected from the surface of the reservoir by a researcher in a kayak. Prior to sampling, the kayak was decontaminated via a 10-minute exposure to 10% bleach solution and then rinsed with reverse osmosis water to remove any viable DNA on the surface of the kayak. To minimize the potential for vectoring eDNA among sampling locations within Lawler Pond, samples were collected, immediately upon arriving at each sampling location, from the bow of the kayak at arms-length (~0.5 m). Additionally, to avoid disturbing future sampling locations, samples were collected starting near the Lawler Pond outflow then proceeded along a single zig-zag pattern ending in the southeast corner of reservoir. The location of each sample was recorded with a handheld GPS (Garmin Corp, Lenexa, Kansas, USA). Each water sample (250-mL bottle) was wiped with a 10% bleach solution and immediately placed in a cooler containing ice for transport back to the laboratory.

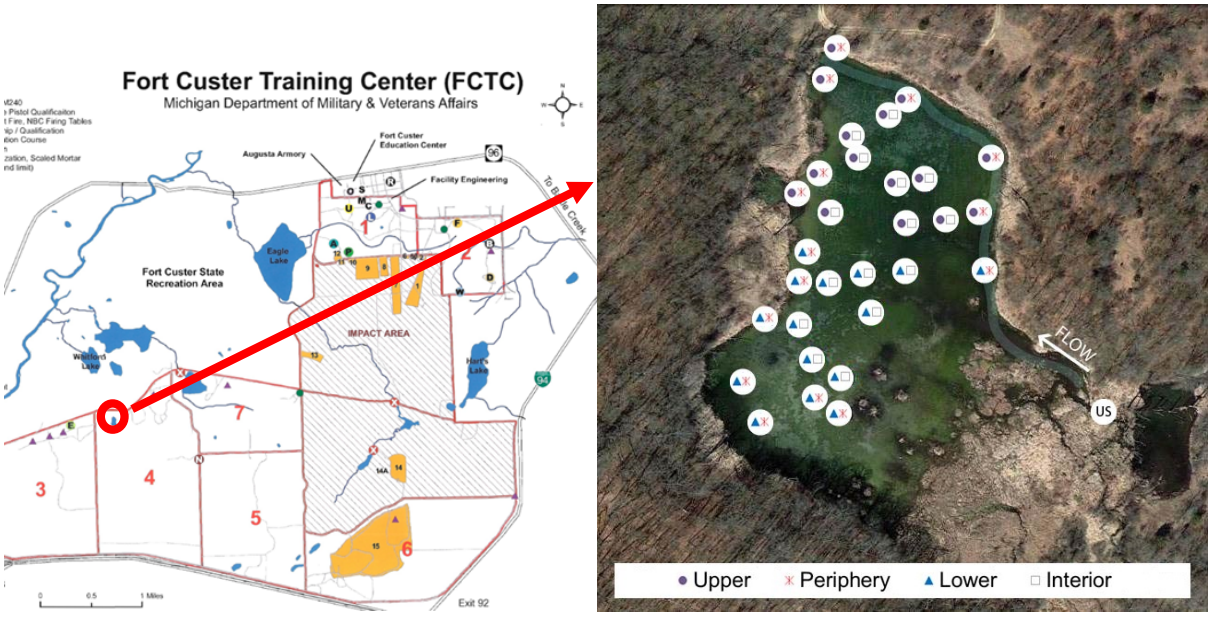


Figure 6. Fort Custer Training facility and aerial photograph of Lawler Pond (Michigan, USA) illustrating the collection location of each eDNA water sample taken from the impoundment and the inflowing stream (US) as well as the location of the deeper channel (shaded). The 15 samples included in each of the four spatial subsampling designs are indicated by the following symbols: circle (upper samples), asterisk (periphery samples), triangle (lower samples), square (interior samples). Each sample was included in two spatial sampling designs as indicated by the two symbols per sample.

Traditional sampling in Lawler Pond used a combination of 17 unbaited metal minnow traps and three unbaited modified-fyke nets, a 2-m diameter cast net, and handheld dip nets. It is important to note that we were not permitted to electrofish in Lawler Pond due to military regulations and safety concerns (i.e., unexploded munitions). Modified-fyke nets were

constructed from two rectangular 91 X 183-cm steel frames, four 76-cm diameter steel hoops, and 13-mm knotless nylon bar mesh. From June 2-6, 2014, all minnow traps and modified-fyke nets were deployed at approximately noon (1200 H), emptied at approximately 1030 H the following morning, then redeployed for a total of four net-nights per net (n=12 total net-nights) and trap (n=68 total trap-nights). Twenty cast net throws were conducted from a boat on the morning of June 6th after the completion of fyke netting. Handheld dip nets were used to target schools of small (<2 cm TL) fishes whenever they were observed. All captured fish were identified to species based on morphological features (and knowledge of local fish fauna), measured for total length and mass, and then returned to the center of the reservoir. In addition to the unbaited minnow traps that are suitable for capturing larval amphibians, surveys were done by sight and by nocturnal call identification. There are 19 species of amphibian known to occur on the Ft. Custer Training Facility property. Included are 12 frog and 7 salamander species. Two of the salamander species are strictly terrestrial and two are present in aquatic environments year-round. Three salamanders are found in the aquatic environment as adults during the breeding season and as larvae during a brief period in the spring and early summer. All of the twelve frog species can be identified by sight and eight can be readily identified by their characteristic vocalizations during the breeding season. From June 6th to 7th 2014 we conducted diurnal sight surveys along with minnow trapping. On two separate nights, one early in March to coincide with breeding period of many of the local frogs and one at the time of intensive trapping in June, we conducted 3-hour long sessions combining sight surveys with call identification.

Environmental DNA was extracted, sequenced, and bioinformatically analyzed according to the methods described in Task 1 except we only evaluated three of the six primer sets in Table 2 (Ac16s, Am12s and L14735/H15149). Species assignment followed that of Figure 4, with the exception that reference list of species used for the USEARCH were those known from Fort Custer along with many related species for which sequence data was available (Table 5).

Table 5. List of species included in the in-house reference sequence database for Lawler pond. Reference sequences taken from previously existing GenBank records are highlighted in blue; reference sequences generated in-house are highlighted in green.

Species	**Ac16S**	**Am12S**	**L14735/H15149**
Acris_crepitans	AY843559	AY843559	
Acris_crepitans_blanchardi			EF988145
Amboplites_rupestris	KM282459	KM282394	KM523260
Ambystoma_laterale	NC006330	NC006330	NC006330
Ambystoma_maculatum			KM523263
Ambystoma_tigrinum	NC006887	NC006887	NC006887
Ameiurus_natalis	AY458872		AY184265
Amphiprion_ocellaris	NC009065	NC009065	NC009065
Bufo_americanus	AY680206	AY680206	AF171190
Bufo_fowleri	AY680224	AY680224	
Catostomus_commersonii	KM282461	KM282400	KM523268
Centropyge_bispinosa	NC028287	NC028287	NC028287
Cottus_bairdii	KM282462	KM282401	KM523269
Cyprinus_carpio	KM282467	KM282406	KM523272

Ecsenius_bicolor	NC028295	NC028295	NC028295
Erimyzon_sucetta	KM282468	KM282408	KM523274
Esox_americanus_vermiculatus			AY497430
Etheostoma_caeruleum	KM282469	KM282409	KM523275
Etheostoma_exile	KM282471	KM282411	KM523277
Etheostoma_nigrum	KM282474	KM282412	KM523280
Etheostoma_radiosum	NC005254	NC005254	NC005254
Hemidactylium_scutatum	DQ283120	DQ283120	NC006342
Hyla_chrysoscelis	EF566949	EF566949	
Hyla_versicolor	AY843682	AY843682	AY843928
Lepomis_cyanellus	KM282484	KM282423	KP013087
Lepomis_gibbosus	KM282485	KM282424	KM523290
Lepomis_gulosus	AY742526		
Lepomis_macrochirus	KM282486	KM282426	KM523292
Lepomis_megalotis	AY742533		AY828977
Lepomis_microlophus	AY742535	*	JF742834
Macropharyngodon_negrosensis	NC028289	NC028289	NC028289
Micropterus_dolomieu	NC011361	KM282429	KM523294
Micropterus_salmoides	KM282489	KM282430	KM523295
Necturus_maculosus			DQ283412
Necturus_maculosus_maculosus	KM282431	KM523296	
Notemigonus_crysoleucus	KM282490	KM282432	KM523297
Notophthalmus_viridescens	EU880323	EU880323	EU880323
Notropis_anogenus			KF744334
Notropis_heterodon	KM282491	KM282434	KM523298
Notropis_stramineus	KM282492	NC008110	KM523299
Oncorhynchus_mykiss	KM282499	KM282441	KM523306
Perca_flavescens	KM282501	KM282443	KM523308
Phoxinus_eos	NC015364	NC015364	NC015364
Pimephales_notatus	AY216556	AY216556	U66606
Pimephales_promelas	KM282503	KM282445	KM523310
Plethodon_cinereus_cinereus	NC006343	NC006343	NC006343
Pomoxis_nigromaculatus	AY742557	KM282446	KM523311
Pseudacris_crucifer			AY210883
Pseudacris_crucifer_crucifer	AY843735	AY843735	
Pseudacris_triseriata	AY843738	AY843738	KJ536224
Pseudanthias_dispar	NC028286	NC028286	NC028286
Rana_catesbeiana	KM282504	NC022696	KM523312
Rana_clamitans	KM282506	DQ283185	KM523314
Rana_palustris	AY779228		
Rana_pipiens	DQ283123	DQ283123	
Rana_sylvatica	DQ283387	DQ283387	AY083271

Rhinichthys_atratulus	AF038495		
Rhinichthys_obtusus	KM282509	KM282447	JX442984
Salarias_fasciatus	AP004451	AP004451	AP004451
Salmo_trutta	KM282510	KM282448	KM523316
Semotilus_atromaculatus	KM282512	AF023199	KM523318
Umbra_limi	KM282516	KM282453	KM523322
Umbra_pygmaea	NC022456	NC022456	NC022456

* This sequence was originally misidentified (by us) as *Lepomis microphus* and then updated as *Lepomis gibbosus.*

Following species assignment, we assessed potential contamination, on a per marker basis, by screening for the presence of any of species detected in the 31 Lawler Pond samples in the mock community, extraction blank, and PCR blank control libraries. If sequence reads from any species were detected in one of the three control libraries, we applied a threshold correction (Hänfling *et al.* 2016; Valentini *et al.* 2016). For the correction, the cumulative relative frequency of contaminant reads for the detected species in the control libraries functioned as a minimum detection threshold, below which a species would not be considered detected. For the Lawler Pond samples, any species with a frequency of occurrence (relative proportion of reads) less than that of the detection threshold were discarded. This correction is like the procedure performed by Hänfling *et al.* (2016), but is based on the false positive reads found in the negative control samples rather than false positive reads found from their mock community species being detected in their field samples.

In addition to controlling for contamination, the effect of bioinformatic decisions on the ability to infer the presence of fishes in Lawler Pond were evaluated using three stringency scenarios representing low, moderate, and high stringency (Figure 7). For the low stringency scenario, a species was considered detected in Lawler pond if its eDNA was found in at least one sample using at least one marker. For the moderate stringency scenario species detection in Lawler Pond required sequences in at least two samples or by at least two markers from a single sample. For the high stringency scenario, sequences from a species were required to be detected in both a minimum of two samples and by a minimum of two markers (species were not required to be detected by the same two markers among samples).

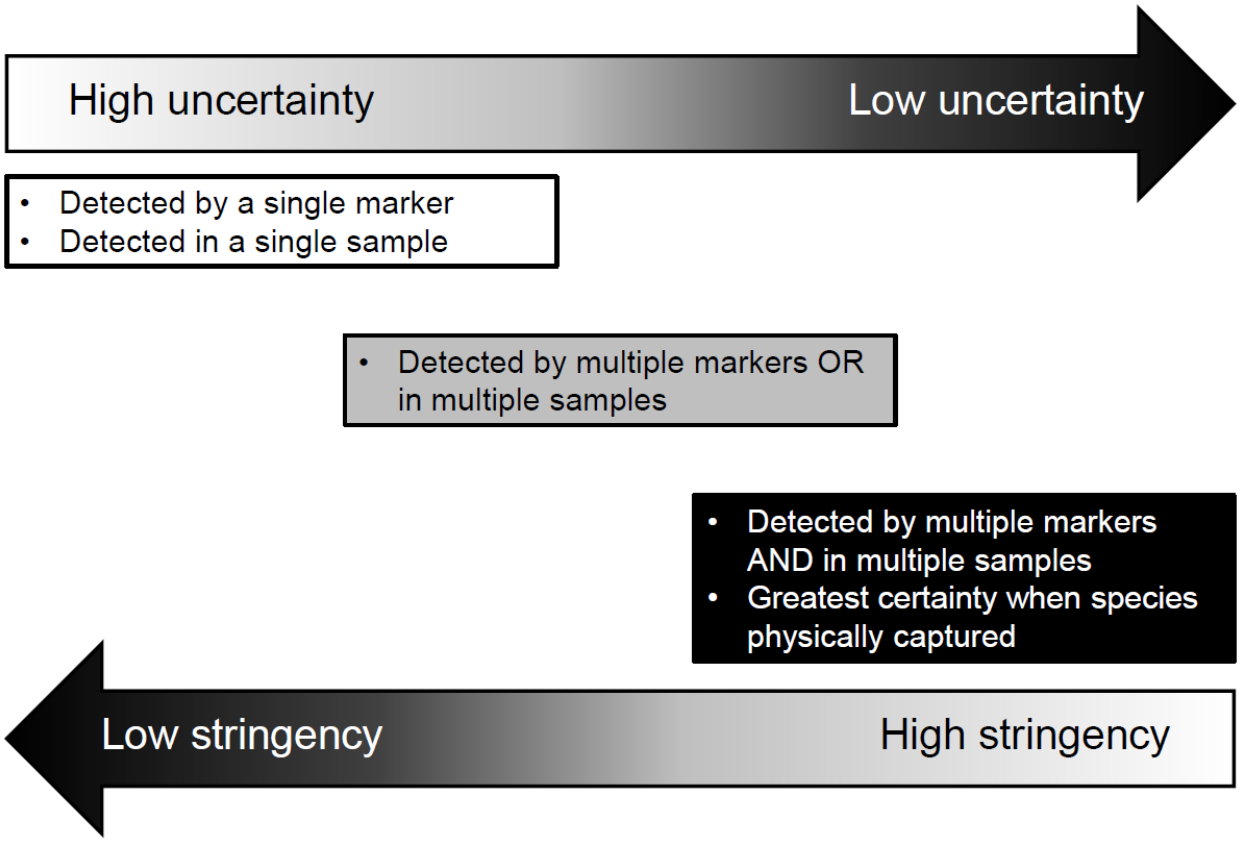


Figure 7. Conceptual diagram illustrating the relationship between bioinformatic stringency and strength of certainty about the presence of eDNA metagenetic-detected species.

Results and discussion

Site 1. Juday Creek

In total, the 240-m reach of Juday Creek surveyed with electrofishing, yielded a total of 12 fish species in 2013, the 17th year of electrofishing in Juday Creek. When combined with historical electrofishing records, the document fish community consists of 18 fish species (Table 6).

Table 6. Historical record of species captured for all reaches within Juday Creek from 1997 - 2013.

Common Name	Scientific Name	# of years captured (of 17 total years)
Creek chub	*Semotilus atromaculatus*	17
Mottled sculpin	*Cottus bairdii*	17
Johnny darter	*Etheostoma nigrum*	17
Western blacknose dace	*Rhinichthys obtusus*	17
White sucker	*Catostomus commersonii*	17
Green sunfish	*Lepomis cyanellus*	17
Rainbow trout	*Oncorhynchus mykiss*	15
Brown trout	*Salmo trutta*	13
Rock bass	*Ambloplites rupestris*	11
Smallmouth bass	*Micropterus dolomieu*	11

Bluegill sunfish	*Lepomis macrochirus*	10
Largemouth bass	*Micropterus salmoides*	6
Pumpkinseed sunfish	*Lepomis gibbosus*	4
Yellow perch	*Perca flavescens*	3
Central mudminnow	*Umbra limi*	2
Rainbow darter	*Etheostoma caeruleum*	2
Warmouth	*Lepomis gulosus*	1
Golden shiner	*Notemigonus crysoleucas*	1

Using our metagenetics approach in 2013, we detected eDNA from all 12 species caught with electrofishing in 2013 and detected four additional species (Table 7). Three of these species had not previously been detected using electrofishing in the 17 years of sampling, yellow bullhead (*Ameiurus natalis*), common carp (*Cyprinus carpio*) and eastern mudminnow (*Umbra pygmaea*), but are known from the region.

Table 7. Species detection indicated by an 'x' for electrofishing and metagenetic methods by stream section ordered from the most downstream section (R1) to the most upstream section (R4). Species detection from eDNA is defined as positive detection by at least two of the four markers (L14735/H15149c, Ac12s, Am12s and Ac16s) in any single sample. D and U represent the downstream and upstream ends of each section. Scientific names given in Table 1, except for yellow bullhead (Ameiurus natalis), common carp (Cyprinus carpio) and eastern mudminnow (Umbra pygmaea).

	Electrofishing				**eDNA metagenetic**							
Common Name	**R1**	**R2**	**R3**	**R4**	**R1 D**	**R1 U**	**R2 D**	**R2 U**	**R3 D**	**R3 U**	**R4 D**	**R4 U**
Rock bass		x	x	x	x	x	x	x	x	x	x	x
Yellow bullhead					x	x	x	x	x	x	x	x
White sucker	x	x	x	x	x	x	x	x	x	x	x	x
Mottled sculpin	x	x	x	x	x	x	x	x	x	x	x	x
Common carp					x	x	x	x	x	x	x	x
Rainbow darter				x		x						
Johnny darter	x	x	x	x	x	x	x	x	x	x	x	x
Green sunfish	x	x	x	x	x	x	x	x	x	x	x	x
Bluegill sunfish	x					x	x	x	x	x	x	x
Smallmouth bass	x	x	x	x	x	x	x	x	x	x	x	x
Largemouth bass					x	x	x	x	x	x	x	x
Rainbow trout		x			x	x	x	x	x	x	x	x
Western blacknose dace	x		x	x	x		x	x	x	x	x	x
Brown trout	x				x		x	x				
Creek chub	x	x	x	x	x	x	x	x	x	x	x	x
Eastern mudminnow					x							

A strong possibility is that some of the eDNA sampled originated from species that occur upstream from where sampling took place. The four reaches sampled using electrofishing were an aggregate 240 m in length, spanning about 750 m of stream, but DNA can persist in the

environment for days (Dejean *et al.* 2011) and can be transported meters to kilometers away from its source (Deiner & Altermatt 2014). Therefore, transport of eDNA into an area where the species are locally absent is a plausible explanation for why they are not detected with the traditional method. Every environment and set of conditions will have different DNA degradation rates (Barnes & Turner 2016) and changing flow dynamics in the case of streams. The detection of carp eDNA in these samples is likely due to the presence of carp or koi (i.e., domesticated common carp) in backyard ponds physically connected to Juday Creek at upstream locations. Additionally, carp are known to inhabit the St. Joseph River into which Juday Creek drains. While suitable carp habitat is not in the sampling areas, suitable habitat does exist upstream, including large in-channel ponds that could support common carp and other species preferring slow-moving water, as well as that for all the non-captured species detected by eDNA. Thus, this study is insufficient to distinguish between the possibility that some species were present in the stream sections, but not captured by electrofishing or the possibility that the species detected occurred only in parts of the watershed upstream from our sampled sections.

Site 2. Lawler pond

In Lawler Pond at Fort Custer trapping yielded a total of ten fish species in 2014 (Table 8). Using a metagenetics approach we detected eDNA from all ten trapped species and as many as 11 additional fish species depending on our level of bioinformatics stringency (Table 8). Historical fish survey records do not exist for Lawler Pond, but records were available for other water bodies at Fort Custer. Those records described a 28-species fish assemblage that includes five of the species detected only by eDNA in 2014 (Shirey *et al.* 2016). Five of the other six species detected only by eDNA have not previously been documented at Fort Custer to our knowledge, but are known to occur within the region. We found that the bioinformatic criteria used to establish enough evidence of a species presence in a sample did change our estimates of richness and use of the moderate level stringency created the least number of false negatives relative to likely false positives when compared to the species caught with traditional sampling (Table 8).

Table 8. Species observed (trapping-based) and detected (eDNA) in Lawler Pond, Fort Custer Training Center, Michigan, under each of the three bioinformatic stringency scenarios: low stringency (Low), moderate stringency (Moderate), and high stringency (High). Black blocks indicate species detected via traditional sampling and/or eDNA metagenetics. Gray blocks indicate eDNA metagenetic false negative detections (i.e., species captured via traditional sampling but not detected with eDNA). White blocks indicate species not detected with either traditional sampling or eDNA metagenetics.

	Traditional	eDNA metagenetic		
Species	Trapping-based	Low	Moderate	High
American Pickerel (*Esox americanus*)	X	X	X	X
Blackchin Shiner (*Notropis heterodon*)	X	X	X	
Bluegill Sunfish (*Lepomis macrochirus*)	X	X	X	X
Bluntnose Minnow (*Pimephales notatus*)		X	X	
Brook Trout (*Salvelinus fontinalis*)		X		
Brown Trout (*Salmo trutta*)		X		

Central Mudminnow (*Umbra limi*)	X	X	X	X
Channel Catfish (*Ictalurus punctatus*)		X		
Common Carp (*Cyprinus carpio*)	X	X	X	X
Creek Chub (*Semotilus atromaculatus*)		X	X	
Green Sunfish (*Lepomis cyanellus*)	X	X	X	
Iowa Darter (*Etheostoma exile*)		X	X	X
Johnny Darter (*Etheostoma nigrum*)		X		
Lake Chubsucker (*Erimyzon sucetta*)		X	X	
Largemouth Bass (*Micropterus salmoides*)	X	X	X	X
Least Darter (*Etheostoma microperca*)		X	X	
Mottled Sculpin (*Cottus bairdii*)		X		
Pumpkinseed Sunfish (*Lepomis gibbosus*)	X	X	X	X
Warmouth Sunfish (*Lepomis gulosus*)	X	X	X	X
White Sucker (*Catostomus commersonii*)		X		
Yellow Bullhead (*Ameiurus natalis*)	X	X	X	
Cumulative Species Richness	***10***	***21***	***15***	***8***

Amphibian surveys detected five species of frogs adjacent to Lawler pond. No frogs or tadpoles we captured by minnow traps. No salamander species were captured or observed in either the terrestrial environment adjacent to the pond or in the aquatic environment. The frog species detected included Blanchard's Cricket Frog (*Acris crepitans*), American Toad (*Bufo americanus*), Eastern Gray Tree frog (*Hyla versicolor*), Spring Peeper (*Pseudacris crucifer*), and the Western Corus Frog (*Pseudacris triserata*). Our eDNA surveys did not detect any amphibian species. Positive detections by call characteristics were largely occurred during the early sampling period in March. Two species (Blanchard's Cricket Frog and Eastern Gray Tree Frog) were detected in the June survey. Two factors may have influenced the lack of amphibian eDNA detections. Since many of the candidate species only use the aquatic environment intermittently, the efficacy of eDNA may be restricted to relatively narrow temporal windows during breeding season or periods of larval development. This pattern is in contract to the constant occupancy for fish. It may be that while we detected five species of frog by their vocalizations they were not utilizing the aquatic environment at the time of our surveys. An additional consideration is that the design and efficiency of our PCR amplifications is biased toward fish species, which may have decreased the probability of a positive eDNA detection for amphibians.

Results from the studies in Task 3 confirm that eDNA metagenetics can detect our benchmark of at least 70% of the fish species in small natural communities. In fact, at both our sites, 100% of fishes found with traditional methods were detected from eDNA, as well as additional species not found with traditional sampling. We recommend that continued use of negative and positive controls is invaluable for assessing potential contaminations that can occur throughout the workflow as demonstrated here. Additionally, based on our tests of bioinformatic filtering (Figure 7), these decisions impact the list of species that can be generated from an eDNA metagenetic method. Therefore, transparency reporting on how many markers and how many samples are used to establish enough evidence for a species presence in a water body is recommended in management applications.

Task 4) Apply the assays to natural environments with unknown species richness at different spatial scales, for comparison with traditional methods for estimating species richness

Methods

Site 3. Eagle Creek, Michigan Army National Guard, Kalamazoo County, Michigan, USA
We collected eDNA samples from nine locations along a longitudinal gradient in Eagle Creek, a tributary to the Kalamazoo River, located within Fort Custer. Additionally, we sampled two locations in the Kalamazoo River directly upstream and downstream of the Eagle Creek confluence (Figure 8). We collected 3 250-mL water samples at each sampling location for a total of 33 eDNA water samples. We also measured the flow velocity at each sampling site to estimate the discharge rate.

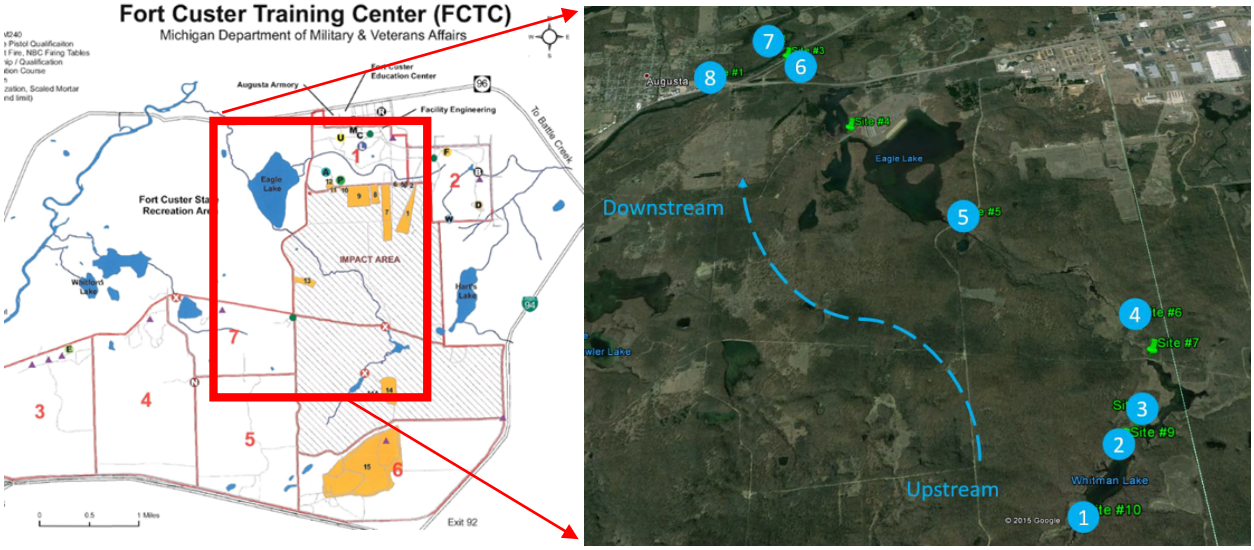


Figure 8. Locality of water samples in Eagle Creek. Site 1 is the most upstream sampling location and Site 8 is the most downstream sampling location. Site 1-6 are all in Eagle Creek. Site 7 and 8 are both in Kalamazoo River.

Filtered water samples were processed, Illumina sequenced, and bioinformatically filtered using the workflow detailed under Task 1. Because we did not have a reference database for the fish species in Eagle Creek, we assigned a species name with each OTU using the NCBI NR database with SAP v1.9.3 (Munch *et al.* 2008) and only used OTUs in further analysis based on assignments to the species level from this method. Contamination from the workflow was removed based on the percentage correction method described for Lawler pond in Task 3 and detailed in Evans *et al.* (2017). Specifically, replicates of the mock community (rep1, rep2) and the PCR negative control were used for the contamination removal of Sample 1-4 (Figure 8). Mock community replicates (rep3, rep4) and the corresponding PCR negative control were used for the contamination removal of Sample 5-8 (Figure 8). Two samples (sample 8a and 8b) were removed from downstream analysis because mock community species reads were detected after percentage correction method in Am12s data.

To infer the presence of fishes in Eagle Creek, we applied the moderate stringency for species detection as outlined in Task 3 (Evans *et al.* 2017), requiring at least two replicates or at least two markers to have a positive detection of each species from a single sampling location.

Site 4. Marine Corps Base Camp Pendleton (MCBCP), California USA

We had two goals for application of the metagenetic approach at MCBCP. First, we wanted to estimate species richness across as many sites as possible and compare this with the known biodiversity from traditional sampling methods. For this analysis, we focus on vertebrates and invertebrates. Second, one of the focal vertebrate species identified by the staff of the Environmental Conservation Division and our research group, was the newly established Southern Tidewater Goby (*Eucyclogobius kristinae*, Swift, Spies, Ellingson, Jacobs 2016). This species is largely restricted in home range to three locations on Camp Pendleton (Swift *et al.* 2016), and with continuing drought conditions at the time of sampling, it was considered a high priority for our surveillance efforts using environmental DNA. For this goal we focused on answering four questions: 1) Where do we detect Tidewater Goby?, 2) How confident are we in our detections using eDNA?, and 3) Are our detection results consistent with more traditions monitoring efforts?

With the aid and guidance of staff from Environmental Conservation Division at MCBCP, we collected 92 250ml water samples across eight locations (Figure 9). Many of these sites were chosen because they are believed to currently or historically provide habitat for Southern Tidewater Goby. All locations are on MCBCP property or have part of their watershed on MCBCP. The Santa Margarita River and Estuary, the largest waterbody sampled, is an approximately 50 km long, intermittent river that drains approximately 1,900 km^2 of watershed. All other sampled locations have smaller watersheds and the sampling effort was focused at the estuaries or lagoons near the confluence with the Pacific Ocean. Tidewater Gobies have historically, been observed in all sampling locations, but as of 2015, physical captured occurred only in San Onofre Creek, Hidden Creek, and Cockleburr Canyon (Swift *et al.* 1989; Swift *et al.* 2016). Water sample collection for eDNA analysis was conducted on April 13 to April 17, 2015.

Filtered water samples were processed, Illumina sequenced, and bioinformatically filtered using the workflow detailed under Task 1. From our metagenetic tool box developed for fish and amphibians, we used the primers sets Ve16S and Ac12S. In addition, we sequenced a 312 bp region of the Cytochrome oxidase I (COI). Primers amplifying a fragment (313 bp) of the COI gene using the MICOIintF (Leray *et al.* 2013) and jgHCO2198 (Geller *et al.* 2013) were shown to capture a large biodiversity for metazoan eukaryotes (Leray *et al.* 2013) and we therefore thought it an advantage to test these primers even though they were not originally tested in our metagenetics toolbox (Table 2). The PCR conditions followed that of Leray *et al.* (2013) and library preparation was the same as the other markers as described in Task 1 and Appendix C2.

We did not have a local reference database for the large targeted group of vertebrate species on Camp Pendleton, we therefore assigned species names in one of two ways. First, we used SAP v1.9.3 (Munch *et al.* 2008) to assign taxa using the NCBI NR database (with a minimum identity of 90%). Second, we used the BLAST function in Geneious v9.1.5 (http://www.geneious.com) against the NCBI NR database using the low complexity and human repeats Filter and used the Entrez Query with the script "all[FILTER] NOT(environmental samples[orgn])" to remove sequences with origins from environmental samples. When SAP and Geneious disagreed, we used the SAP assignment. After taxonomic assignments, we observed many species had a pattern

of one highly abundant OTU with several low abundant OTUs. Using our mock community species as a guild, we further reduced the OTUs with assignments by requiring the NCBI BLAST hit to have a coverage with our OTU > 99 % and a sequence similarity > 98%.

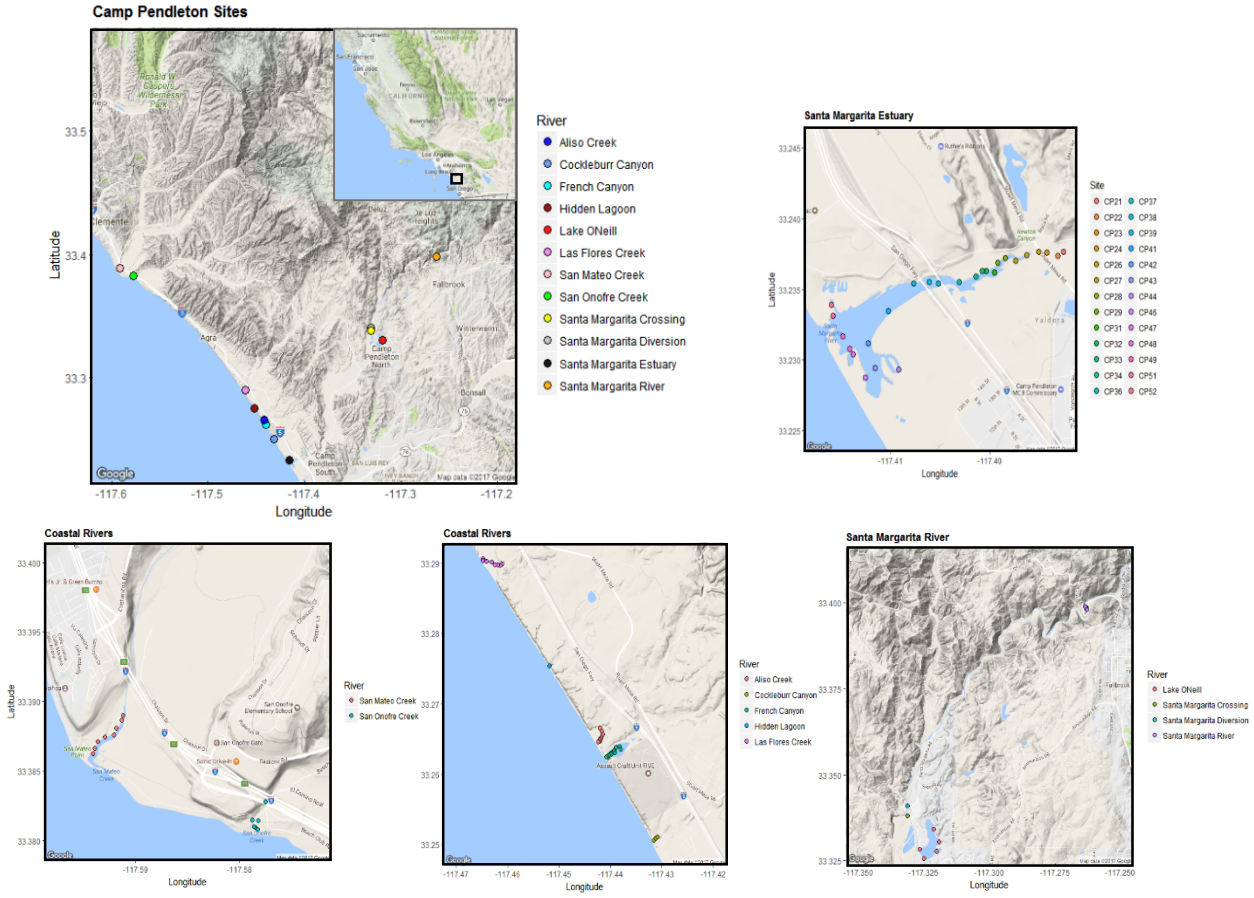


Figure 9. Locations from where water samples were collected on or near MCB Camp Pendleton. Large map shows the twelve locations that were sampled and smaller areas show sampling effort within each location.

Additionally, because of the study goal of identifying the presence of the Southern Tidewater Goby, we introduced a new level of negative controls, 40 in total. The negative controls consisted of cooler blanks (*n*=12), which were 250 mL bottles filled with distilled water, that were carried into the field with samples. Each location had at least one cooler blank randomly assigned in the collection sequence with additional water blanks in the Santa Margarita River due to the larger sampling effort (Figure 9). The remaining controls were laboratory quality controls. Specifically, we generated eight extraction blanks, eight PCR blanks, four index blanks, and eight mock community samples. Extraction blanks were constructed by using only extraction reagents without a filter and subsequently processed alongside field samples and other controls in the lab. PCR blanks were no-template controls (NTC) applied to each of the three gene regions amplified. The NTC consisted of sterile water that was added as template during the first round of PCR amplification. A band was then excised from the agarose gel at the anticipated amplicon size, cleaned, and used as template for the second round of PCR amplification, which included the addition of a unique barcode. Index blanks are used in the last step of library preparation and evaluates if samples could become contaminated if many libraries in the past

were prepared with the same Illumina adapters. A single mock community was constructed and run in parallel from the DNA extraction step with eight replicates. The mock community was composed of equal amounts of tissue derived DNA (measured with Qubit dsDNA HS Assay, Life Technologies, Carlsbad, California, USA) from six species of fish endemic to Southeastern US and were not expected in the study area (*Umbra limi, Thoburnia atripinnis, Erimyzon sucetta, Notropis topeka,* and *Noturus taylori*).

With our 40 negative controls, we described the distribution of errant sequence reads for each mitochondrial fragment for the Southern Tidewater Goby (a.k.a. background contamination). In Juday Creek (Olds *et al.* 2016), four negative controls were used to generate error distributions and evaluate the probability for each field sample, that the observed number of sample reads (or greater) could come from the error distribution by chance (Appendix A2). Since the data are represented in counts of positive DNA strands per sample, we can assume a Poisson distribution describes the number of errant strands we should expect to see in any given sample. The maximum likelihood estimator (MLE) for the mean of the Poisson distribution is $\hat{\lambda} = \frac{\sum x_i}{n}$. Error distribution estimation followed protocols previously described in Task 3.

To compare our metagenetic eDNA results with that from historic data for species with known distributions within and nearby MCBCP, we compiled lists of species that have previously been detected on or near MCBCP with the assistance of data available from the Environmental Conservation Division. We additionally merged this data with data from the Global Biodiversity Information Facility's occurrence database (http://www.gbif.org/occurrence). We downloaded all occurrence data from the bounding box designated by the polygon with the following coordinates (-117.64 33.21,-117.64 33.44,-117.21 33.44,-117.21 33.21,-117.64 33.21) on May 25, 2017. To infer the presence of species detected, we applied the moderate stringency for species detection as outlined in Task 3 (Evans et al. 2017), requiring at least two replicates from a site or at least two markers to consider the species as a positive detection in a river.

Results and discussion

Site 3. Eagle Creek

In Eagle Creek, eDNA metagenetic analysis detected 23 species, which were all detected before by traditional sampling in Kalamazoo River by the MI DNR between 1968 and 2005 (Figure 10).

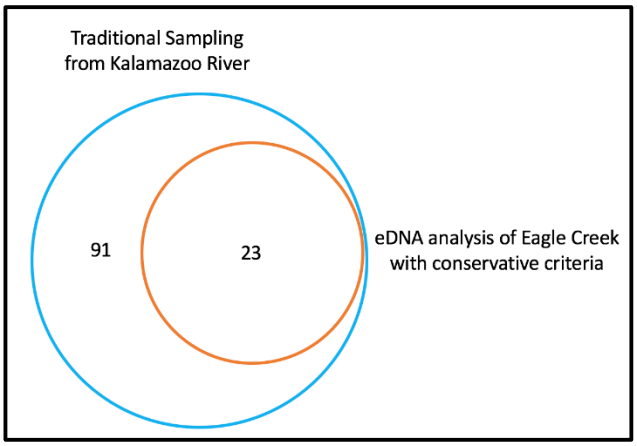


Figure 10. Venn diagram of shared species detections between traditional sampling of Kalamazoo River watershed between 1968 and 2005 by the MI DNR with eDNA data in Eagle Creek with moderate species calling criteria.

The number of detected fish species increased from upstream to downstream (Table 9), with only two fish species detected (*Umbra limi* and *Esox americanus*) at the furthest upstream location and 10 species detected at the mouth of Eagle Creek. Nine and 11 species were found in the two Kalamazoo River sampling locations (Site 7 and 8).

Table 9. Species found in each locality in Eagle Creek and Kalamazoo River.

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8
Lepomis_macrochirus								
Ameiurus_natalis								
Lepomis_gibbosus								
Semotilus_atromaculatus								
Micropterus_salmoides								
Etheostoma_nigrum								
Ambloplites_rupestris								
Moxostoma_erythrurum								
Etheostoma_caeruleum								
Umbra_limi								
Perca_flavescens								
Notemigonus_crysoleucas								
Hypentelium_nigricans								
Etheostoma_exile								
Amia_calva								
Percina_maculata								
Noturus_flavus								
Micropterus_dolomieu								
Lepomis_gulosus								
Lepomis_cyanellus								
Esox_americanus								
Erimyzon_sucetta								
Cyprinella_spiloptera								
Total	2	7	5	8	8	10	9	11

Unlike Juday Creek in task 3, results from this lotic system show that eDNA transport may not have a strong influence on eDNA metagenetic detections downstream. For example, we found some fish species (e.g., *Micropterus salmoides*) in an upstream site (site 2), but not in an immediate downstream site (site 3), and we detected it again at site 4 further downstream. It is possible that DNA transport distance is low given the low discharge rate of in Eagle Creek (114 L/s).

At the catchment scale for Eagle creek, we successfully detected many species expected to occur from very little sampling effort and we therefore conclude that the metagenetic method for species detection at this broad scale will be important for tracking general species richness for the watershed and can be used for sampling areas where intense electrofishing or other traditional methods are not permitted for safety reasons (e.g., unexploded ordinances, etc.). However, further research is needed to determine the transport of eDNA in natural lotic systems to determine the accuracy in detection on a more localized scale.

Site 4. Camp Pendleton

Goal 1: Species richness compared to historical knowledge

We detected 37 species using an eDNA metagenetic approach. Twenty-seven native species were detected on or near the base (Table 10). Of the four listed native species the southern

steelhead was the only species not detected. We additionally detected ten exotic species all previously documented on the base (Table 10). The successful detection of the three important types of biodiversity (native, protected species and exotic) demonstrates the power of this tool to monitor biodiversity of interest on the MCBCP.

However, it is also clear that careful criteria are followed to establish enough evidence for a species presence at a location using the eDNA metagenetic method of species detection. For species of concern (protected or exotic) we recommend contamination control criteria are met and these criteria are outlined through the example below for the Southern Tidewater Goby. We have applied this criteria to the four protected species and the ten exotic species (Appendix A3).

Table 10: Detections of species in watersheds within or near Camp Pendleton. All species have previously been documented on the base. Green rows are native species, blue rows are native species with state or federal protection and orange rows are exotic species. Asterisks by exotic species indicate the species was confirmed present through traditional sampling in 2013. The 24 native, but non-listed species presence has not been controlled for contamination by the same process as the listed or exotic species and will be in the final report.

Common name (*species name*)	San Mateo Creek	San Onofre Creek	Las Flores Creek	Hidden Lagoon	Aliso Canyon	French Canyon	Cockle-burr Canyon	Santa Margarita Estuary	Upper Santa Margarita River	Santa Margarita Crossing/ Diversion	Lake O'Neill
Deep body anchovy (*Anchoa compressa*)	0	0	0	0	0	0	0	1	0	0	0
Topsmelt silverside (*Atherinops affinis*)	0	0	0	0	1	0	0	1	0	0	0
Arrow goby (*Clevelandia ios*)	0	0	0	0	0	1	0	1	0	0	0
California killifish (*Fundulus parvipinnis*)	0	1	0	1	1	0	0	1	1	0	0
Longjaw mudsucker (*Gillichthys mirabilis*)	1	1	1	1	1	1	1	1	0	1	1
Opaleye (*Girella nigricans*)	0	0	0	0	0	0	0	1	0	0	0
California butterfly ray (*Gymnura marmorata*)	0	0	0	0	0	0	0	1	0	0	0
Diamond turbot (*Hypsopsetta guttulata*)	0	0	0	0	0	0	0	1	0	0	0
Pacific staghorn sculpin (*Leptocottus armatus*)	0	0	0	0	0	0	0	1	0	0	0
California corbina (*Menticirrhus undulatus*)	0	0	1	0	1	0	0	1	0	0	0
Southern vole (*Microtus levis*)	1	0	0	0	0	0	0	0	0	0	0
Striped mullet (*Mugil cephalus*)	0	0	1	0	0	0	0	1	0	1	1
Grey smooth-hound shark (*Mustelus californicus*)	0	0	0	0	0	0	0	1	0	0	0
Bat ray (*Myliobatis californica*)	0	0	0	0	0	0	0	1	0	0	0
Spotted sand bass (*Paralabrax maculatofasciatus*)	1	0	0	0	0	0	0	1	0	0	0
Barred sand bass (*Paralabrax nebulifer*)	1	0	0	0	0	0	0	1	0	0	0
California halibut (*Paralichthys californicus*)	0	0	0	0	0	0	0	1	0	0	0
Pacific tree frog (*Pseudacris regilla*)	0	1	0	0	0	0	0	0	0	1	0
Shovelnose guitarfish (*Rhinobatos productus*)	0	0	0	0	0	0	0	1	0	0	0

Spotfin croaker (*Roncador stearnsii*)	1	0	0	0	0	1	0	1	0	0	0
Pacific Sardine (*Sardinops sagax*)	1	0	0	0	0	0	0	0	0	1	1
California sheephead (*Semicossyphus pulcher*)	1	0	0	0	0	0	0	0	0	0	0
Yellowfin croaker (*Umbrina roncador*)	0	0	0	0	0	0	0	1	0	0	0
North American beaver (*Castor canadensis*)	1	0	1	0	0	0	0	1	1	1	0
Arroyo toad (*Anaxyrus californicus*)	0	0	0	0	0	0	0	0	1	0	0
Arroyo chub (*Gila orcuttii*)	0	0	0	0	0	0	0	0	0	1	0
Tidewater goby *(Eucyclogobius newberryi* (now *E. kristinae*))	0	1	0	1	0	1	1	0	0	0	0
Steelhead (*Oncorhynchus mykiss*)	0	0	0	0	0	0	0	0	0	0	0
Yellowfin goby (*Acanthogobius flavimanus*)	0	0	0	0	0	0	0	*1	0	0	0
Black bullhead (*Ameiurus melas*)	*1	0	0	0	0	0	0	0	0	0	0
Brown bullhead (*Ameiurus nebulosus*)	1	1	0	0	0	0	0	0	1	0	0
Common carp (*Cyprinus carpio*)	0	0	0	0	0	0	1	0	*1	0	1
Western mosquitofish (*Gambusia affinis*)	0	1	0	0	0	0	1	0	0	0	0
Green sunfish (*Lepomis cyanellus*)	*1	1	0	0	0	0	0	0	*1	0	0
Redeye bass (*Micropterus coosae*)	0	0	0	0	0	0	0	0	*1	0	0
Largemouth bass (*Micropterus salmoides*)	0	0	0	0	0	0	0	0	*1	0	1
Black crappie (*Pomoxis nigromaculatus*)	0	0	0	0	0	0	0	0	0	0	1
American bullfrog (*Lithobates catesbeiana*)	*1	*1	0	0	0	0	0	0	*1	*1	1

Goal 2: Assessment of metagenetic method for detection of the Southern Tidewater Goby
In the 40 negative control samples, there was background contamination, as expected (Olds *et al.* 2016). The MLE of the Poisson distribution for each of the gene fragments showed variation in average background DNA; AC12s, Vert16s, and COI had 0.9, 10.03, and 0.25 strands of DNA respectively. Consequently, the cutoff thresholds of number of DNA strands to ensure only a 0.001 probability of the observed number of DNA strands occurred by chance was different for each gene fragment; AC12s, Vert16s, and COI had thresholds of 6, 21, and 1, respectively. Using these thresholds, we dismissed many observations with trace amounts of DNA as detection due to background DNA contamination. Once accounted for however, we had a number of positive detections.

Of the eight locations sampled, we had positive detections in multiple gene fragments at San Onofre Creek, Hidden Lagoon, Cockleburr Canyon, and French Canyon. We had positive detections in one gene fragment in the Santa Margareta River and Estuary, and no detections at San Mateo Creek, Las Flores, Aliso Canyon (Table 11).

Table 11. Detections of Southern Tidewater Goby on and near Camp Pendleton, California.

Locations	Goby status as of 2015 (Swift et al. 2016)	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
				AC12s	Vert16s	COI
San Mateo Creek	Historically present	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
San Onofre Creek	Recently collected	Detected (15 April 2015)	6	6 (100%)	6 (100%)	5 (83.3%)
Las Flores Creek	Historically collected	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Recently collected	Detected (13 April 2015)	4	3 (75%)	3 (75%)	3 (75%)
Aliso Canyon	Historically present	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
French Canyon	Historically present	Detected (14 April 2015)	12	8 (66.7%)	9 (75%)	3 (25%)

Cockleburr Canyon	Recently collected	Detected (13 April 2015)	4	1 (25%)	2 (50%)	2 (50%)
Upper Santa Margarita River	Unknown	Inconclusive (13 April 2015)	8	0 (0%)	7 (88%)	0 (0%)
Middle Santa Margarita River	Unknown	Not Detected (17 April 2015)	3	0 (0%)	0 (0%)	0 (0%)
Santa Margarita Estuary	Historically present	Not Detected (13 &14 April 2015)	26	0 (0%)	1 (3.8%)	0 (0%)

Using an environmental metagenetic method for the detection of the Southern Tidewater Goby holds great promise for a reliable, non-invasive survey method. Using Swift *et al.*'s (2016) paper as our best evaluation of current Southern Tidewater Goby distribution in estuaries using traditional methods, we have one location with a discrepancy between approaches. Traditional gear did not capture Southern Tidewater Goby in French Creek, but as Lafferty (2013) pointed out, an upstream refugia may provide a source of recolonization of the sampled estuary. This upstream refugia may also provide eDNA into the estuary, thus explaining the absence of traditional detection, but positive eDNA detection. DNA was recovered in seven samples of one gene fragment of the Upper Santa Margareta River. If the condition that at least two gene fragments indicate presence is required, we would then consider the eDNA evidence to be inconclusive about the presence of the Southern Tidewater Goby at this location. There are no concurrent traditional sampling records at this location to evaluate this detection, but further investigation seems warranted.

The removal of background contamination is critical for assessing species presence and absence from eDNA (Olds *et al.* 2016). The negative controls revealed that trace DNA can infiltrate the process and we cannot determine at which point during the collection, processing, and/or sequencing of samples. In our study, we saw very few errant reads particularly in the AC12s and COI gene fragments, yet because negative controls had reads, we can describe the error distribution and samples with six and one or fewer reads can be dismissed as errant detections. The Vert16s gene fragment had more errant reads than the other fragments requiring a threshold of 21 reads. Empirically, from the negative controls we know that without accounting for error distributions, the false positive probability is 33/120 =0.28 but after correction the false positive probability is 3/120=0.025. With replication at each location and the criteria that at least two of the gene fragments or two samples amplified eDNA for the species at each location to confirm a positive detection further decreases the probability of false positives.

Task 5) Apply and compare alternative models for estimating species richness and biodiversity across a variety of aquatic systems

Methods

Estimation of species richness (number of species) is based on detection probabilities (Boulinier *et al.* 1998), which often vary by species, habitat, and detection method. Accurate estimates of species richness are difficult to obtain because detection probabilities are generally much lower than 100%, particularly for rare or elusive species in aquatic habitats. Low detection probabilities cause severe underestimates of species richness (Smith & van Belle 1984) and various mathematical approaches can be applied to correct for this underestimation (Hellmann & Fowler 1999).The most widely-used and best-performing of these estimators all rely on the fundamental principle that the abundances of the rarest species (or their frequencies in a set of samples) can be used to estimate the frequencies of undetected species (Gotelli & Colwell 2011).Two of the most common approaches to correct for undetected species are the Chao I and Chao II estimators (Gotelli & Colwell 2011).

The bias-corrected Chao I estimator,

$$S_{ChaoI} = S_{obs} + \frac{f_1(f_1 - 1)}{2(f_2 + 1)}$$

is applied to *abundance data* (where the abundance of each species in a sample is recorded). S_{obs} is the number of species observed, f_1 is the number of species observed only once (singletons), and f_2 is the number of species observed only twice (doubletons).

The bias-corrected Chao II estimator,

$$S_{Chao2} = S_{obs} + \frac{n-1}{n} * \frac{q_1(q_1 - 1)}{2(q_2 + 1)}$$

is applied to *incidence data* (where only the presence of each species in a sample is recorded). S_{obs} is the number of species observed, n is the sample size, q_1 is the number of species present in only one sample, and q_2 is the number of species present in only two samples. The estimator includes (n-1)/n to adjust for small sample sizes, which is needed when samples sizes are small.

We hypothesized that the eDNA metagenetic approach would result in more accurate estimates of species richness by increasing detection probability for multiple species. To test this hypothesis, we used data from the mesocosm experiment in Task 2, and the natural systems evaluated in Task 3: Juday Creek and Lawler Pond. We calculated species richness for both systems with the bias-corrected Chao II estimator.

Results and discussion

The hypothesis for task five was that eDNA estimation of species richness would perform as well, or better than, traditional gears, such as electroshocking, nets, and traps. We had three unique studies to draw upon. First, the mesocosm replicates had a total of nine species manipulated at various densities (Table 3). In all mesocosm replicates, eDNA detected all nine species. There was no estimation as we knew the manipulated densities and there was no variation in the eDNA observed species richness, so no error bars surround the species richness value (Figure 11). Second, Juday Creek had consistent expected species richness, using Chao estimators (~16 species), between eDNA and traditional gear methods both efforts were sufficient to quantify the uncertainty around the estimates (Figure 11). Third, Lawler Pond, had

sufficient eDNA effort to create a species richness estimate, but did not have sufficient effort to create a species richness estimate for traditional gears (no error bars for x-axis in Figure 11). Nevertheless, in all studies, eDNA performed as well or better than traditional capture. While admittedly, our three studies do not hold sufficient statistical power to make a definitive conclusion about the eDNA metagenetic approach's ability to quantify species richness, there is a rapidly growing body of literature that is finding similar patterns of eDNA species detection outperforming more traditional sampling efforts (Andruszkiewicz *et al.* 2017; Lim *et al.* 2016; Port *et al.* 2016; Shelton *et al.* 2016; Sigsgaard *et al.* 2016; Stoeckle *et al.* 2017; Thomsen *et al.* 2016).

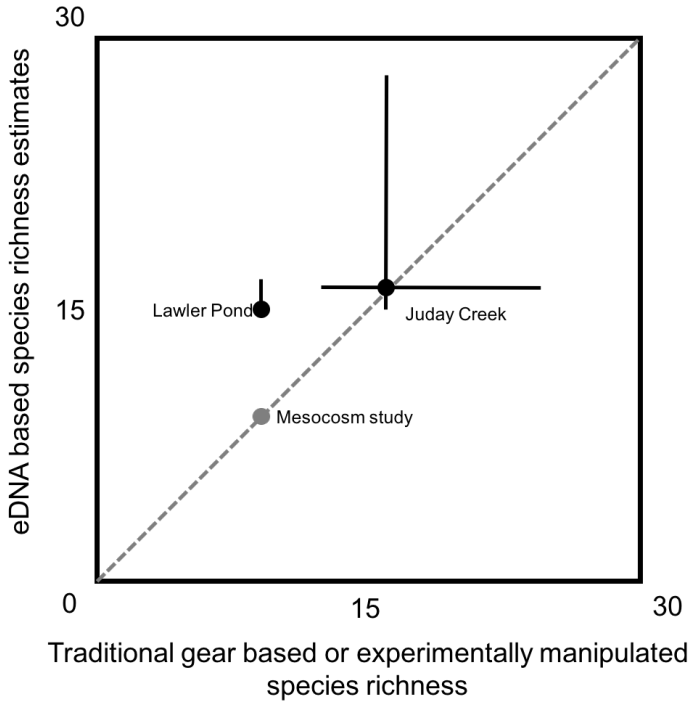


Figure 11: Species richness estimated from gear based capture or experimentally manipulated communities. Gray circle indicates the mesocosm study where species richness was exactly known. Black circles indicate the natural study systems, where species richness was estimated with the two types of sampling tools. Error bars represent 95% confidence intervals using a Chao II estimator. Dashed grey line indicates the one to one relationship.

Unexpected outcomes and synergies

At the onset of our research, the application of a metagenetic approach to surveying fish and amphibian biodiversity was at a nascent stage. During our research and that of others in the field (Deiner *et al.* 2016; Hänfling *et al.* 2016; Valentini *et al.* 2016) it became clear that this method has immense promise for the future of assessing community richness for animal species. Because the discoveries of what is possible with this genetic tool are rapidly evolving, we were able to advanced research for many unexpected outcomes due to synergies that were only possible because of the ongoing research with this project.

Specifically, we could leverage our unique samples from the mesocosm experiment and Juday Creek and in-house funds to empirically demonstrate that eDNA can be in the genomic state for at least the mitochondria (approximately 16 to 17 kilo bases in length) from fishes currently inhabiting a water body. We tested a method of using long-range PCR coupled with Illumina sequencing of eDNA and recovered mitochondrial genomes for ten of the twelve species caught

with traditional methods. We additionally sequenced whole mitochondrial genomes from the two species we detected only from the eDNA in our previous study (Olds *et al.* 2016). This research is transformative for the field of eDNA research because it tested an important assumption, namely that eDNA is degraded and therefore the laboratory strategy must focus on short amplicon fragments (100-400 bp) for sequencing. Focus on such short fragments limits the ability to assign species level information to sequences obtained from environmental samples because there are not always enough sequence differences between closely related species to confidently assign a species name to a sequence. A manuscript of our methods and findings is under consideration with the journal of *Methods in Ecology and Evolution* (Deiner *et al.* In press).

An additional unexpected outcome came from when the SERDP scientific committee directed our group to evaluate the transport of environmental DNA in streams based on our white paper. Using partial support from the SERDP program and additional funding secured by Co-PI Jerde through the University of Notre Dame's Environmental Change Initiative internal grants program ($74,339). The group, comprised of Dr. Jennifer Tank (Ecology), Dr. Diogo Bolster (Engineering), and Dr. Brett Olds (SERDP supported post-doc) along with Arial Shogren (Ph.D. graduate student with Tank) and Notre Dame undergraduate student Elizabeth Andruszkiewicz (Supervised by Bolster) worked with the SERDP group to produce three papers detailing their findings (Jerde et al. 2016, Shogren et al. 2016, and Shogren et al. In revision).

In Jerde *et al.* (2016), we conducted a series of seminatural stream experiments to test the sensitivity of new digital droplet PCR (ddPCR) to detect low concentrations of eDNA in a lotic system, measure the residence time of eDNA compared to a conservative tracer, and we model the transport of eDNA in this system. We found that while ddPCR improves our sensitivity of detection, the residence time and transport of eDNA does not follow the same dynamics as the conservative tracer and necessitates a more stochastic framework for modeling eDNA transport. There was no evidence for differences in the transport of eDNA due to substrate type. The relatively large amount of unexplained variability in eDNA transport revealed the need for uncovering mechanisms and processes by which eDNA is transported downstream leading to species detections, particularly when inferences are to be made in natural systems where eDNA is being used for conservation management.

In Shogren *et al.* (2016) we conducted experiments in continuous flow columns packed with porous substrates to explore eDNA transport dynamics and asked whether substrate type and the presence of colonized biofilms plays a role for eDNA retention. To interpret our data, and for modelling purposes, we began with the assumption that eDNA could be treated as a classical tracer. Comparing our experimental data with traditional transport models, we found that eDNA behaves anomalously, displaying characteristics of a heterogeneous, polydisperse substance with particle-like behavior that can be filtered by the substrate. Columns were quickly flushed of suspended eDNA particles while a significant number of particles never made it through and were retained in the column, as calculated from a mass balance. Suspended eDNA was exported through the column, regardless of biofilm colonization. Our results indicate that the variable particle size of eDNA results in stochastic retention, release and transport, which may influence the interpretation eDNA detection in biological systems.

Lastly in Shogren *et al.* (In revision) we used an empirical approach and a simple conceptual model to propose a framework of how eDNA is transported, retained, and resuspended in stream

systems. Such an understanding of eDNA dispersal in streams is essential for designing optimized sampling protocols and subsequently estimating biomass or organismal abundance. We also provide guiding principles for more effective use of eDNA methods, highlighting the necessity of understanding these parameters for use in predictive modeling of eDNA transport.

Conclusions and Implications for Future Research/Implementation

In this project, we contributed dramatic progress to the understanding that eDNA metagenetic methods have very high sensitivity and accuracy of in estimating species richness relative to traditional methods for sampling fishes. Results from this project and other simultaneous projects also identified some important frontiers of eDNA research with important management implications. We highlight here three questions ripe for progress.

First, what are the rates of production and loss (transport, sedimentation, degradation) of eDNA for different kinds of organisms and under different environmental conditions? Without more quantification of production and loss rates, we cannot confidently interpret the time of production and/or distance from the sampling site of an organism whose eDNA is detected.

Second, what is the relationship under different environmental conditions between eDNA read number and the population size of the species detected? Management actions could be more reliably designed if population size, in addition to presence, were known.

Third, given our discovery that whole mitochondrial genomes can be sequenced from eDNA, to what extent could mitochondrial (or even nuclear) genetic variation be determined from eDNA to inform population genetics studies? Genetic variation is an important aspect of population viability for many species.

Literature cited

Andruszkiewicz EA, Starks HA, Chavez FP*, et al.* (2017) Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. *PLOS ONE* **12**, e0176343.

Barnes MA, Turner CR (2016) The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics* **17**, 1-17.

Beja- Pereira A, Oliveira R, Alves PC, Schwartz MK, Luikart G (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources* **9**, 1279-1301.

Benton N, Ripley J, Powledge F (2008) Conserving biodiversity on military lands: A guide for natural resources managers. NatureServe, Arlington, Virgina.

Boulinier T, Nichols JD, Sauer JR, Hines JE, Pollock K (1998) Estimating species richness: the importance of heterogeneity in species detectability. *Ecology* **79**, 1018-1028.

Burgener M, Hübner P (1998) Mitochondrial DNA enrichment for species identification and evolutionary analysis. *Zeitschrift für Lebensmitteluntersuchung und-Forschung A* **207**, 261-263.

Creer S, Fonseca VG, Porazinska DL*, et al.* (2011) Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Mol Ecol* **19**.

Deiner K, Altermatt F (2014) Transport distance of invertebrate environmental DNA in a natural river. *PLOS ONE* **9**, e88786.

Deiner K, Fronhofer EA, Mächler E, Walser J-C, Altermatt F (2016) Environmental DNA reveals that rivers are conveyer belts of biodiversity information. *Nature Communications* **7**.

Deiner K, Renshaw MA, Li Y*, et al.* (In press) Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. *Methods in Ecology and Evolution*.

Dejean T, Valentini A, Duparc A*, et al.* (2011) Persistence of environmental DNA in freshwater ecosystems. *PLOS ONE* **6**, e23398.

Dudgeon D (2010) Prospects for sustaining freshwater biodiversity in the 21st century: linking ecosystem structure and function. *Current Opinion in Environmental Sustainability* **2**, 422-430.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461.

Evans NT, Li Y, Renshaw MA*, et al.* (2017) Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*.

Evans NT, Olds BP, Renshaw MA*, et al.* (2016) Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources* **16**, 29-41.

Geller J, Meyer C, Parker M, Hawk H (2013) Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all- taxa biotic surveys. *Molecular Ecology Resources* **13**, 851-861.

Goldberg CS, Turner CR, Deiner K*, et al.* (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*.

Gotelli NJ, Colwell RK (2011) Estimating species richness. In: *Biological diversity: frontiers in measurement and assessment* (eds. Magurran A, McGill B), pp. 39-54. Oxford University Press, Oxford, UK.

Hänfling B, Lawson Handley L, Read DS*, et al.* (2016) Environmental DNA metabarcoding of lake fish communities reflects long- term data from established survey methods. *Molecular Ecology*.
Hellmann JJ, Fowler GW (1999) Bias, precision, and accuracy of four measures of species richness. *Ecological applications* **9**, 824-834.
Jerde CL, Mahon AR, Chadderton WL, Lodge DM (2011) "Sight- unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters* **4**, 150-157.
Jerde CL, Olds BP, Shogren AJ*, et al.* (2016) Influence of Stream Bottom Substrate on Retention and Transport of Vertebrate Environmental DNA. *Environmental science & technology* **50**, 8770-8779.
Kitano T, Umetsu K, Tian W, Osawa M (2007) Two universal primer sets for species identification among vertebrates. *International journal of legal medicine* **121**, 423-427.
Leray M, Yang JY, Meyer CP*, et al.* (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in zoology* **10**, 34.
Lim NK, Tay YC, Srivathsan A*, et al.* (2016) Next-generation freshwater bioassessment: eDNA metabarcoding with a conserved metazoan primer reveals species-rich and reservoir-specific communities. *Royal Society Open Science* **3**, 160635.
Lodge DM, Turner CR, Jerde CL*, et al.* (2012) Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology* **21**, 2555-2558.
Lodge DM, Williams S, MacIsaac HJ*, et al.* (2006) Biological invasions: recommendations for US policy and management. *Ecological applications* **16**, 2035-2054.
Magnuson JJ, Benson BJ, McLain AS (1994) Insights on species richness and turnover from long-term ecological research: fishes in north temperate lakes. *American Zoologist* **34**, 437-451.
Magurran AE (2010) Q&A: What is biodiversity? *BMC biology* **8**, 145.
Munch K, Boomsma W, Huelsenbeck JP, Willerslev E, Nielsen R (2008) Statistical assignment of DNA sequences using Bayesian phylogenetics. *Systematic Biology* **57**, 750-757.
Olds BP, Jerde CL, Renshaw MA*, et al.* (2016) Estimating species richness using environmental DNA. *Ecology and Evolution*.
Pilliod DS, Goldberg CS, Arkle RS, Waits LP (2013) Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences* **70**, 1123-1130.
Port JA, O'Donnell JL, Romero- Maraccini OC*, et al.* (2016) Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology* **25**, 527-541.
Riaz T, Shehzad W, Viari A*, et al.* (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, gkr732.
Sala OE, van Vuuren D, Pereira HM*, et al.* (2005) Biodiversity across scenarios. In: *Ecosystems and human well-being: scenarios: findings of the scenarios working group* eds. Carpenter SR, Pingali PL, Bennett EM, Zurek MB), p. 375. Millennium Ecosystem Assessment.
Schmidt BR, Pellet J (2009) Quantifying abundance: counts, detection probabilities, and estimates. *Amphibian Ecology and Conservation: A Handbook of Techniques. Oxford University Press, Oxford*, 465-479.
Shelton AO, O'Donnell JL, Samhouri JF*, et al.* (2016) A framework for inferring biological communities from environmental DNA. *Ecological applications* **26**, 1645-1659.
Shirey P, Brueseke M, Kenny J, Lamberti G (2016) Long-term fish community response to a reach-scale stream restoration. *Ecology and Society* **21**.

Shogren AJ, Tank JL, Andruszkiewicz EA*, et al.* (2016) Modelling the transport of environmental DNA through a porous substrate using continuous flow-through column experiments. *Journal of The Royal Society Interface* **13**, 20160290.
Shogren AJ, Tank JL, Andruszkiewicz EA*, et al.* (In revision) Controls on eDNA movement in streams: Transport, Retention, and Resuspension. *Scientific Reports*.
Sigsgaard EE, Nielsen IB, Bach SS*, et al.* (2016) Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution* **1**, 0004.
Smith EP, van Belle G (1984) Nonparametric estimation of species richness. *Biometrics*, 119-129.
Stoeckle MY, Soboleva L, Charlop-Powers Z (2017) Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary. *PLOS ONE* **12**, e0175186.
Swift C, Nelson J, Maslow C, Stein T (1989) Biology and distribution of the tidewater goby, Eucyclogobius newberryi (Pisces: Gobiidae) of California. *Natural History Museum of Los Angeles County* **404**.
Swift CC, Spies B, Ellingson RA, Jacobs DK (2016) A New Species of the Bay Goby Genus Eucyclogobius, Endemic to Southern California: Evolution, Conservation, and Decline. *PLOS ONE* **11**, e0158543.
Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* **21**, 2045-2050.
Takahara T, Minamoto T, Doi H (2013) Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLOS ONE* **8**, e56584.
Thompson W (2013) Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters, pp. 1-7. Island Press, New York.
Thomsen PF, Møller PR, Sigsgaard EE*, et al.* (2016) Environmental DNA from Seawater Samples Correlate with Trawl Catches of Subarctic, Deepwater Fishes. *PLOS ONE* **11**, e0165252.
Turner CR, Barnes MA, Xu CC*, et al.* (2014) Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution* **5**, 676-684.
Valentini A, Taberlet P, Miaud C*, et al.* (2016) Next- generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology* **25**, 929–942.
Wheeler TJ, Eddy SR (2013) nhmmer: DNA homology search with profile HMMs. *Bioinformatics*, btt403.
Xenopoulos MA, Lodge DM, Alcamo J*, et al.* (2005) Scenarios of freshwater fish extinctions from climate change and water withdrawal. *Global Change Biology* **11**, 1557-1564.

Appendix A.

All raw data and supplemental files associated with all publications are freely available and all publications have been published in open access. We provide here the DOI links to all papers and data associated with each task. One supplemental file from Olds et al. 2016 has been added below (A2) because it played a particular important role for interpreting data in Task 4.

Task 1:
eDNA pipeline code is available from the website https://github.com/pfrender-laboratory/epps
Renshaw et al. 2014: http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12281/abstract

Task 2:
All raw data associated with this study have been deposited on the NCBI's Sequence Read Archive (SRA) (www.ncbi.nlm.nih.gov/sra) under the BioProject PRJNA317862
Evans et al. 2016: http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12433/full

Task 3:
All raw data associated with these studies have been deposited on the NCBI's Sequence Read Archive (SRA) (www.ncbi.nlm.nih.gov/sra) under the BioProject PRJNA317862
Olds et al. 2016: http://onlinelibrary.wiley.com/doi/10.1002/ece3.2186/full
Evans et al. 2017: https://doi.org/10.1139/cjfas-2016-0306

Task 4:
All raw data associated with these studies have been deposited on the NCBI's Sequence Read Archive (SRA) (www.ncbi.nlm.nih.gov/sra) under the BioProject PRJNA317862

Task 5:
Olds et al. 2016: http://onlinelibrary.wiley.com/doi/10.1002/ece3.2186/full

A1. Reference data base built for assigning species names with sequences obtained from metagenetic analysis of water samples. The reason a species was listed is because it is known to occur in one of three locations: TE (from threatened or endangered list), ME (from mesocosm species list), JC (from Juday Creek species list), and FC (from Fort Custer species list). A species in red indicates no tissues were available to generate primary sequence data for the metagenetic primers. Red cells for a metagenetic primer set means no sequence data is available. Orange cells for a metagenetic primer set indicate the sequences were used from previously reported public data from NCBIs GenBank and the accession number is listed. Grey cells are all sequence we generated in-house and are publically available from NCBIs GenBank and the accession number is listed.

Taxa	Reason	Scientific name	Common name	Am12S	Ac12S	L14735/H15149c (CYTB)	Ac16S	HL16S	Vert16S
Fish	TE	*Acipenser brevirostrum*	Shortnose sturgeon	KM282389	KM273793	KM523255	KM282454	KM434916	KM434992
Fish	TE	*Acipenser medirostris*	Green sturgeon	KM282390	KM273794	KM523256	KM282455	KM434917	KM434993
Fish	TE	*Acipenser oxyrinchus*	Atlantic sturgeon	KM282391	KM273795	KM523257	KM282456	KM434918	KM434994
Fish	TE	*Acipenser oxyrinchus desotoi*	Gulf sturgeon	KM282392	KM273796	KM523258	KM282457	KM434919	KM434995
Fish	TE	*Alosa alabamae*	Alabama shad	KM282393	KM273798	KM523259	KM282458	KM434921	KM434996
Fish	JC	*Ambloplites rupestris*	rock bass	KM282394	KM273799	KM523260	KM282459	KM434922	KM434997
Fish	TE	*Amblyopsis rosae*	Ozark cavefish						
Fish	FC	*Ameiurus natalis*	yellow bullhead		DQ421865	AY184265	AY458872	AY458872	AY458872
Fish	FC	*Ameiurus nebulosus*	brown bullhead		AY458873	AY458889	AY458873	AY458873	AY458873
Fish	FC	*Amia calva*	bowfin	NC_004742	NC_004742	NC_004742	NC_004742	NC_004742	NC_004742
Fish	ME	*Campostoma anomalum*	central stoneroller	KM282399	KM273807	KM523267	KM282460	KM434929	KM435001
Fish	ME,JC, FC	*Catostomus commersonii*	white sucker	KM282400	KM273808	KM523268	KM282461	KM434930	KM435002
Fish	JC	*Cottus bairdii*	mottled sculpin	KM282401	KM273809	KM523269	KM282462	KM434931	KM435003
Fish	JC	*Cottus cognatus*	slimy sculpin	AB188190	AB188190	AY116365			KJ778623
Fish	TE	*Cyprinella callitaenia*	Bluestripe shiner	KM282402	KM273810		KM282463	KM434932	KM435004
Fish	TE	*Cyprinella zanema pop 2*	Coastal plain Santee chub	KM282403	KM273811	KM523270	KM282464	KM434933	KM435005
Fish	TE	*Cyprinodon macularius*	Desert pupfish	KM282404	KM273812	KM523271	KM282465	KM434934	KM435006

Fish	TE	*Cyprinodon tularosa*	White sands pupfish	KM282405	KM273813	NC_028292	KM282466	KM434935	KM435007
Fish	ME,FC	*Cyprinus carpio*	common carp	KM282406	KM273814	KM523272	KM282467	KM434936	KM435008
Fish	TE	*Elassoma okatie*	Bluebarred pygmy sunfish	KM282407	KM273815	KM523273		KM434937	KM435009
Fish	FC	*Erimyzon oblongus*	creek chubsucker	NC_013064	NC_013064	NC_013064	NC_013064	NC_013064	NC_013064
Fish	FC	*Erimyzon sucetta*	lake chubsucker	KM282408	KM273816	KM523274	KM282468	KM434938	KM435010
Fish	FC	*Esox americanus vermiculatus*	grass pickerel			AY497430			
Fish	FC	*Esox lucius*	northern pike	NC_004593	NC_004593	NC_004593	NC_004593	NC_004593	NC_004593
Fish	JC	*Etheostoma caeruleum*	rainbow darter	KM282409	KM273817	KM523275	KM282469	KM434939	
Fish	TE	*Etheostoma cragini*	Arkansas darter	KM282410	KM273818	KM523276	KM282470	KM434940	KM435011
Fish	TE	*Etheostoma ditrema*	Coldwater darter			FJ012507			
Fish	FC	*Etheostoma exile*	Iowa darter	KM282411	KM273819	KM523277	KM282471	KM434941	KM435012
Fish	TE	*Etheostoma maculatum*	Spotted darter			KM523278	KM282472	KM434942	KM435013
Fish	TE	*Etheostoma microlepidum*	Smallscale darter		KM273820	KM523279	KM282473	KM434943	KM435014
Fish	JC	*Etheostoma nigrum*	Johnny darter	KM282412	KM273821	KM523280	KM282474	KM434944	KM435015
Fish	TE	*Etheostoma okaloosae*	Okaloosa darter		KM273822	JF314767	KM282475	KM434945	
Fish	TE	*Etheostoma sellare*	Maryland darter						
Fish	TE	*Etheostoma tuscumbia*	Tuscumbia darter	KM282413	KM273823	KM523281	KM282476	KM434946	KM435016
Fish	TE	*Eucyclogobius newberryi*	Tidewater goby	KM282414	KM273824	KM523282	KM282477	KM434947	KM435017
Fish	ME	*Fundulus notatus*	blackstripe topminnow	KM282416	KM273826	KM523285	KM282478	KM434950	KM435019
Fish	ME	*Gambusia holbrooki*	Eastern mosquitofish	KM282417	KM273827	KM523286	KM282479	KM434951	KM435020
Fish	TE	*Gasterosteus aculeatus williamsoni*	Unarmored threespine stickleback	KM282418	KM273828	KM523287	KM282480	KM434952	KM435021
Fish	TE	*Gila bicolor mohavensis*	Mohave tui chub	KM282419	KM273829	KM523288	KM282481	KM434953	KM435022
Fish	TE	*Gila orcutti*	Arroyo chub	KM282420	KM273830		KM282482	KM434954	KM435023
Fish	TE	*Iotichthys phlegethontis*	Least chub	KM282422	KM273832	KM523289	KM282483	KM434956	KM435025
Fish	FC	*Lepisosteus oculatus*	spotted gar	NC_004744	NC_004744	NC_004744	NC_004744	NC_004744	NC_004744
Fish	JC,FC	*Lepomis cyanellus*	green sunfish	KM282423	KM273833	KP013087	KM282484	KM434957	KM435026

Fish	JC,FC	*Lepomis gibbosus*	pumpkinseed	KM282424	KM273834	KM523290	KM282485	KM434958	KM435027
Fish	JC,FC	*Lepomis (Chaenobryttus) gulosus*	warmouth	KM282425	KM273835	KM523291	AY742526	AY742526	AY742526
Fish	JC,FC	*Lepomis macrochirus*	bluegill	KM282426	KM273836	KM523292	KM282486	KM434959	KM435028
Fish	JC	*Lepomis microlophus*	redear sunfish			JF742834	AY742535	AY742535	AY742535
Fish	TE	*Meda fulgida*	Spikedace	KM282427	KM273837	NC_028291	KM282487	KM434960	KM435029
Fish	TE	*Microphis brachyurus (lineatus)*	Opossum pipefish	KM282428	KM273838	KM523293	KM282488	KM434961	KM435030
Fish	JC	*Micropterus dolomieu*	smallmouth bass	KM282429	KM273839	KM523294	NC_011361	KM434962	KM435031
Fish	JC	*Micropterus punctulatus*	spotted bass	KM370980	KM370980	HM070928	AY742548	AY742548	AY742548
Fish	JC,FC	*Micropterus salmoides*	largemouth bass	KM282430	KM273840	KM523295	KM282489	KM434963	KM435032
Fish	JC,FC	*Notemigonus crysoleucas*	golden shiner	KM282432	KM273842	KM523297	KM282490	KM434964	KM435033
Fish	FC	*Notropis anogenus*	pugnose shiner			KF744334			
Fish	FC	*Notropis chalybaeus*	ironcolor			KC763697			
Fish	FC	*Notropis heterodon*	blackchin shiner	KM282434	KM273844	KM523298	KM282491	KM434966	KM435035
Fish	FC	*Notropis heterolepis*	blacknose shiner			AY140696			
Fish	FC	*Notropis stramineus*	sand shiner	NC_008110	NC_008110	KM523299	KM282492	KM434967	KM435036
Fish	FC	*Notropis texanus*	weed shiner	AY216552	AY216552	AF352267			
Fish	TE	*Notropis topeka*	Topeka shiner	KM282435	KM273844	KM523300	KM282493	KM434968	KM435037
Fish	FC	*Noturus gyrinus*	tadpole madtom	AY015534	AY015534	AY327295	AY458874	AY458874	AY458874
Fish	TE	*Noturus taylori*	Caddo madtom	KM282436	KM273846	KM523301	KM282494	KM434969	KM435038
Fish	TE	*Oncorhynchus clarkii henshawi*	Lahontan cutthroat trout	KM282437	KM273847	KM523302	KM282495	KM434970	KM435039
Fish	TE	*Oncorhynchus clarkii stomias*	Greenback cutthroat trout	KM282438	KM273848	KM523303	KM282496	KM434971	KM435040
Fish	TE	*Oncorhynchus keta*	Chum salmon	KM282439	KM273849	KM523304	KM282497	KM434972	KM435041
Fish	TE	*Oncorhynchus kisutch*	Coho salmon	KM282440	KM273850	KM523305	KM282498	KM434973	KM435042
Fish	TE,JC	*Oncorhynchus mykiss*	Steelhead trout	KM282441	KM273851	KM523306	KM282499	KM434974	KM435043
Fish	TE	*Oncorhynchus tshawytscha*	Chinook salmon	KM282442	KM273852	KM523307	KM282500	KM434975	KM435044
Fish	JC,FC	*Perca flavescens*	yellow perch	KM282443	KM273853	KM523308	KM282501	KM434976	KM435045
Fish	TE	*Percina cymatotaenia*	Bluestripe darter			AF386589			

Fish	TE	*Percina rex*	Roanoke logperch	KM282444	KM273854	KM523309	KM282502	KM434977	KM435046
Fish	TE	*Percina sp. 12 (Percina brucethompsoni)*	Ouachita darter			KM209991			
Fish	FC	*Phoxinus eos*	northern redbelly dace	NC_015364	NC_015364	NC_015364	NC_015364	NC_015364	NC_015364
Fish	ME	*Pimephales promelas*	fathead minnow	KM282445	KM273855	KM523310	KM282503	KM434978	KM435047
Fish	FC	*Pomoxis nigromaculatus*	black crappie	KM282446	KM273856	KM523311	NC_028298	NC_028298	NC_028298
Fish	JC	*Rhinichthys atratulus*	blacknose dace	AP012104	AP012104	AP012104	AP012104	AP012104	AP012104
Fish	JC	*Rhinichthys obtusus*	Western blacknose dace	KM282447	AF023198	JX442984	KM282509	KM434984	KM435052
Fish	JC	*Salmo trutta*	brown trout	KM282448	KM273859	KM523316	KM282510	KM434985	KM435053
Fish	TE	*Salvelinus confluentus*	Bull trout	KM282449	KM273860	KM523317	KM282511	KM434986	KM435054
Fish	JC	*Semotilus atromaculatus*	creek chub	AF023199	AF023199	KM523318	KM282512	KM434987	KM435055
Fish	TE	*Thoburnia atripinnis*	Blackfin sucker	KM282451	KM273862	KM523320	KM282514	KM434989	KM435057
Fish	TE	*Tiaroga cobitis*	Loach minnow	KM282452	KM273863	KM523321	KM282515	KM434990	KM435058
Fish	JC,FC	*Umbra limi*	Central mudminnow	KM282453	KM273864	KM523322	KM282516	KM434991	KM435059
Fish	JC	*Umbra pygmaea*	Eastern mudminnow	AP013049	AP013049	AP013049	AP013049	AP013049	AP013049
Amphibian	FC	*Acris crepitans blanchardi*	Blanchard's cricket frog	AY843559	KM273797	EF988145	AY843559	KM434920	AY843559
Amphibian	TE	*Ambystoma bishopi*	Reticulated flatwood salamander	KM282395	KM273800	KM523261		KM434923	KM434998
Amphibian	TE	*Ambystoma californiense*	California tiger salamander	KM282396	KM273801	KM523262	NC_006890	KM434924	KM434999
Amphibian	TE	*Ambystoma cingulatum*	Frosted flatwood salamander	DQ283184	KM273802	EF036621	DQ283184	KM434925	DQ283184
Amphibian	FC	*Ambystoma laterale*	blue-spotted salamander	NC_006330	NC_006330	NC_006330	NC_006330	NC_006330	NC_006330
Amphibian	FC	*Ambystoma maculatum*	spotted salamander		KM273803	KM523263		KM434926	
Amphibian	FC	*Ambystoma tigrinum*	Eastern tiger salamander	NC_006887	NC_006887	NC_006887	NC_006887	NC_006887	NC_006887
Amphibian	TE	*Ambystoma tigrinum stebbinsi*	Sonoran tiger salamander	KM282397	KM273804	KM523264	KP013120	KM434927	KM435000
Amphibian	TE	*Batrachoseps wrightorum*	Oregon slender salamander	KM282398	KM273805	KM523265	NC_006333	KM434928	NC_006333
Amphibian	FC	*Bufo americanus*	American toad	AY680206	AY680206	AF171190	AY680206	AY680206	AY680206

Amphibian	TE	*Bufo (Anaxyrus) californicus*	Arroyo toad	AY680225	KM273806	KM523266	AY680225	AY680225	AY680225
Amphibian	FC	*Bufo (woodhousii) fowleri*	Fowler's toad	AY680224	AY680224		AY680224	AY680224	AY680224
Amphibian	TE	*Eurycea neotenes*	Texas salamander	KM282415	KM273825	KM523283		KM434948	KM435018
Amphibian	TE	*Eurycea tridentifera*	Comal blind salamander			KM523284		KM434949	
Amphibian	TE	*Gyrinophilus palleucus*	Tenessee cave salamander	KM282421	KM273831	NC_028297	NC_028297	KM434955	KM435024
Amphibian	FC	*Hemidactylium scutatum*	four-toed salamander	DQ283120	DQ283120	NC006342	DQ283120	DQ283120	DQ283120
Amphibian	FC	*Hyla chrysoscelis*	Cope's gray treefrog	EF566949	EF566949		EF566949	EF566949	EF566949
Amphibian	FC	*Hyla versicolor*	Eastern gray treefrog	AY843682	AY843682	AY843928	AY843682	AY843682	AY843682
Amphibian	FC	*Necturus maculosus maculosus*	mudpuppy	KM282431	KM273841	KM523296	DQ283412	DQ283412	DQ283412
Amphibian	TE	*Notophthalmus perstriatus*	Striped newt	KM282433	KM273843	NC_028278	NC_028278	KM434965	KM435034
Amphibian	FC	*Notophthalmus viridescens*	Eastern newt	EU880323	EU880323	EU880323	EU880323	EU880323	EU880323
Amphibian	FC	*Plethodon cinereus cinereus*	red-backed salamander	NC_006343	NC_006343	NC_006343	NC_006343	NC_006343	NC_006343
Amphibian	FC	*Pseudacris crucifer crucifer*	Spring peeper	AY843735	AY843735	AY210883	AY843735	AY843735	AY843735
Amphibian	FC	*Pseudacris triseriata*	Western chorus frog	AY843738	AY843738	KJ536224	AY843738	AY843738	AY843738
Amphibian	ME,FC	*Rana catesbeiana*	American bullfrog	NC_022696	NC_022696	KM523312	KM282504	KM434979	KM435048
Amphibian	FC	*Rana clamitans melanota*	Northern green frog	DQ283185	KM273857	KM523314	KM282506	KM434981	DQ283185
Amphibian	TE	*Rana (aurora) draytonii*	California red-legged frog	NC_028296	NC_028296	NC_028296	KM282507	KM434982	KM435050
Amphibian	TE	*Rana chiricahuensis*	Chiricahua leopard frog	DQ583270	DQ583270	KM523313	KM282505	KM434980	KM435049
Amphibian	TE	*Rana okaloosae*	Florida bog frog	NC_028283	NC_028283	KM523315	KM282508	KM434983	KM435051
Amphibian	FC	*Rana palustris*	pickerel frog		KM273858		AY779228	AY779228	AY779228
Amphibian	FC	*Rana pipiens*	Northern leopard frog	DQ283123	DQ283123		DQ283123	DQ283123	DQ283123
Amphibian	FC	*Rana sylvatica*	wood frog	DQ283387	DQ283387	AY083271	DQ283387	DQ283387	DQ283387

Appendix A2: Estimating contamination rates and assessing false positives for detection of fishes with eDNA in Juday Creek, Indiana, USA. Supplement from Olds et al. 2016.

The use of quality controls is extremely important in high-throughput eDNA metagenetics studies to investigate artifacts such as background contamination and recovery of expected DNA through bioinformatics workflows (Murray et al 2015). In our study of Juday Creek, five such quality controls were included during the laboratory processing and high-throughput sequencing (HTS) of eDNA samples. We included two full process, "cooler blanks" consisted of reverse osmosis water filled in a sampling container in the lab and transported during field collections (one on each of the two sampling days). These cooler blanks where subsequently filtered and extracted along-side samples from Juday Creek and used in PCR through to library preparation and HTS. The third control was a PCR negative control that was used at the first step of PCR and carried through library preparation and HTS. Forth, a positive PCR control that consisted of a pool of DNA extracted from tissues of tropical marine fish not known to occur in the area and were PCR-amplified alongside field collected samples to create a mock community. These four controls were then used to monitor for contamination during the laboratory phase during the study. A fifth control, known as PhiX, was added as a control to the libraries before running on the MiSeq for HTS. This control, and its results, are additionally discussed in the Supplemental Fig. 2 in Olds et al. 2016 and will not be considered further here.

A total of 16 species were identified from field samples using four different markers for a total of 64 marker-species combinations, each having four potential observations that could be observed in controls. If there was no contamination in any negative control, then we would expect to see no sequences from these 16 species in any of the four markers. However, in 46% of control observations (119 of 256), one or more sequences were assigned to a marker-species combination when there should have been zero sequences. On average, there were 18 sequences per marker-species combination across the four controls (n=256, Std. Dev. = 76). There was a significant difference among markers (GLM; Poisson error with log-link function, p<0.01). Actino12s marker averaged 45 sequences per marker-species combination (Std. Dev. = 141), while Actino 16s, Amphibia12s and Cyt B averaged 3 (Std. Dev. = 8), 8 (Std. Dev. = 23), and 14 (Std. Dev. = 42), respectively.

With increasing amounts of sequences observed for a given species in a field sample, it may be expected that there is increased potential for DNA to contaminate a control. To test this hypothesis, we evaluated correlations between the total number sequences observed per species for each marker in field samples with the number of sequences per species for each marker in each of the four control samples. The correlation was significantly positive for all markers: Actino12s (r=0.98, p<0.01), Actino16s (r=0.93, p<0.01), Amphibia12S (r=0.99, p<0.01), and CytB (r=0.99, p<0.01). As such, there is evidence that the error distribution of the amount of contaminant DNA is unique to each marker-species combination and is correlated to the observed number of sequences for that species and marker in the field controls.
Since the data are represented in counts of DNA sequences per sample, we can assume each of the four control observations are from a Poisson distribution that describes the number of errant sequences that we should expect to see in any given quality control. The maximum likelihood estimator for the mean of the Poisson distribution is $\hat{\lambda} = \frac{\sum x_i}{n}$. While other count data distributions may be better suited, such as negative binomial (McMurdie & Holmes 2014), there is little empirical evidence to support using these distributions at this time and with only four data

points, estimating more than one parameter for the error distribution is difficult to justify. We choose to move forward working on the assumption of a Poisson error distribution and recognize the need for more research to better justify this assumption – presumably from a study with more quality controls simultaneously collected and analyzed. Below we walk through this logic applied to an example with the fish Johnny Darter and enumerate this for all other species in Supplemental Table S2 in Olds et al. (2016), to test whether low levels of contamination, no matter their source, can influence the outcome of biological interpretation for the species we detect in Juday Creek. We used a stringent threshold of more than two sequences observed for at least two markers to consider a species present at a field site. In addition, we chose to use a statistical approach rather than the arbitrary removal of sequences from the whole dataset as currently practiced (e.g., Valentini et al. 2015), to ask the question at this threshold what is the likelihood of concluding a species is at a site when it in in fact not.

For example, the Johnny Darter (*Etheostoma nigrum*) had sequences numbers of 11, 4, 81, and 168 (Table S7, Olds et al. 2016) for the Ac12s marker observed in the four quality controls. These counts are the number of sequences from samples that should have no sequences present and therefore represent background levels of contamination during our eDNA assay. The mean number of copies as estimated from the Poisson distribution is $\hat{\lambda} = 66$. There were eight samples collected and processed from Juday Creek and each sample had a number of sequences observed for the Ac12s marker for Johnny Darter. Given our estimated distribution of errant Ac12s Johnny Darter sequences in the quality controls, we can use the Cumulative Mass Function (CMF) of the Poisson distribution to ask, "What is the chance of the error distribution producing the observed number of Johnny Darter sequences in a field sample?" For the Ac12s Johnny Darter detection, this results in Probability of (X>x) <0.0001. The conclusion is that for each field sample, the detection of Johnny Darter in Ac12s is unlikely due to contamination. But there are three additional markers. Applying the same approach, we find that for all samples across all markers there are sufficient Johnny Darter sequences in each sample to conclude the Johnny Darter detection is likely not due to contamination in any field sample (Appendix A2 Table 1).

The same approach was applied to the remaining 15 species listed as detected in Juday Creek (Table S2, Olds et al. 2016). When we account for the possible detection due to contamination, there are 20 instances with some >0.001 possibility that contamination has led to the positive detection for a single species-marker combination. However, with the criteria that more than two sequences had to of been observed in at least two markers, this would only change our interpretation of three observations of the possible 128 (Appendix A2 Table 2). The affected two species the Blue Gill and Yellow Bullhead. For the other species for which we observed a significant value (i.e., Rainbow Trout, Common Carp, and Brown Trout) there was enough evidence from the other markers that we could exclude the marker showing a possible contamination and still have enough sequence evidence to conclude its presence in the sample.

While three species-samples combinations changed from positive detection to no detection, all species are still detected and there is no change to the expected species richness. However, there is the potential that the contamination-corrected species counts for each field site have change the confidence intervals of the Chao estimators. There is very little difference between the Chao estimates and their 95% confidence intervals with and without contamination considered (Appendix A2 Table 3). This is because there was no change in the total observed species richness (16) and the shift from detection to non-detection for the three samples (two samples for

Yellow Bullhead and one sample for Bluegill) occurred for species where the sample incidence record shifted from eight to six and from eight to 7 for Yellow Bullhead and Bluegill respectively. Overall, we conclude that contamination has little influence on our conclusions about the observed or estimated species richness in Juday Creek.

Appendix A2 Table 1: Parameter estimates (λ) and associated probabilities that the observed number of Johnny Darter sequences in field samples came from the distribution of errant DNA.

		Juday Creek samples							
		L1	L2	L3	L4	L5	L6	L7	L8
Marker	Ac12s (λ=66)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Ac16s (λ=6.5)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Am12s (λ=12.25)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Cytb (λ=92.5)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Appendix A2 Table 2: Detection of species as a result of presence of DNA in samples and consideration of contamination. Dark gray cells indicate species and samples that switched from detection to non-detection as a result of considering possible contamination leading to detection.

	Sites in Juday Creek							
	Most downstream ←----------→Most upstream							
Species	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**	**R8**
Johnny Darter	Detection	Detection	Detection	Detection	Detection	Detection	Detection	Detection
Mottled Sculpin	Detection	Detection	Detection	Detection	Detection	Detection	Detection	Detection
White Sucker	Detection	Detection	Detection	Detection	Detection	Detection	Detection	Detection
Green Sunfish	Detection	Detection	Detection	Detection	Detection	Detection	Detection	Detection
Rainbow Trout	Detection	Detection	Detection	Detection	Detection	Detection	Detection	Detection
Creek Chub	Detection	Detection	Detection	Detection	Detection	Detection	Detection	Detection
Rock Bass	Detection	Detection	Detection	Detection	Detection	Detection	Detection	Detection
Bluegill	Non Detection	Detection	Detection	Detection	Detection	Detection	Detection	Detection
Largemouth Bass	Detection	Detection	Detection	Detection	Detection	Detection	Detection	Detection
Smallmouth Bass	Detection	Detection	Detection	Detection	Detection	Detection	Detection	Detection
Common Carp	Detection	Detection	Detection	Detection	Detection	Detection	Detection	Detection
Yellow Bullhead	Detection	Detection	Detection	Detection	Non Detection	Detection	Non Detection	Detection
Western Blacknose Dace	Detection	Non Detection	Non Detection	Non Detection	Detection	Detection	Detection	Detection
Rainbow Darter	Non Detection	Detection	Non Detection	Non Detection	Non Detection	Non Detection	Non Detection	Non Detection
Brown Trout	Detection	Detection	Detection	Detection	Non Detection	Non Detection	Non Detection	Non Detection
Eastern Mudminnow	Detection	Non Detection	Non Detection	Non Detection	Non Detection	Non Detection	Non Detection	Non Detection

Appendix A2 Table 3: Comparison of mean Chao estimator and confidence interval for incidence base accumulation curve without and with contamination considered.

Sample	Without contamination considered Mean (95% CI)	With contamination considered Mean (95% CI)
1	13.74	13.31
2	14.51	14.49
3	14.95	14.92
4	15.33	15.33
5	15.58 (15.39, 20.17)	15.58 (15.4, 20.19)
6	16.12 (15.65, 23.60)	16.12 (15.65, 23.61)
7	16.62 (15.92, 26.86)	16.62 (15.92, 26.87)
8	16.88 (16.06, 28.65)	16.88 (16.06, 28.65)

Appendix A2: References

McMurdie P. J. and Holmes S. (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Computational Biology* 10(4): e1003531.

Murray DC, Coghlan ML, Bunce M. (2015) From benchtop to desktop: important considerations when designing amplicon sequencing workflows. *PloS ONE* 10:e0124671.

Olds BP, Jerde CL, Renshaw MA*, et al.* (2016) Estimating species richness using environmental DNA. *Ecology and Evolution*.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., et al. (2015) Next-generation monitoring of aquatic biodiversity using environmental DNA metagenetics. *Molecular Ecology*. Accepted Author Manuscript. doi:10.1111/mec.13428.

Appendix A3: Tables showing presence of listed and exotic species after contamination control. Our current analysis uses a detection criteria of having at least one positive sample in at least two markers OR at least two positive samples within one marker. If we make the criteria more stringent by having at least one positive detection in at least two markers then there are some noticeable changes to our conclusions. The Tidewater Goby detection in the upper Santa Margarita River is changed to not detected. Similarly, Largemouth Bass presence in the upper Santa Margarita River, Green Sunfish in San Onofre Creek, Brown Bullhead in San Mateo Creek, San Onofre Creek, and Upper Santa Margarita River, and Common Carp in Cockleburr Canyon all change from detected to not detected. The expected outcome is lowered species richness at each of these location, but it may also result in higher false negative rates.

Appendix A3 Table 1 (same as Table 11 in main document): Detections of **Tidewater Goby** on and near Camp Pendleton, California

Locations from Northern to Southern	Goby status as of 2015 (Swift et al. 2016)	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
				AC12s	Vert16s	COI
San Mateo Creek	Historically present	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
San Onofre Creek	Recently collected	Detected (15 April 2015)	6	6 (100%)	6 (100%)	5 (83.3%)
Las Flores Creek	Historically collected	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Recently collected	Detected (13 April 2015)	4	3 (75%)	3 (75%)	3 (75%)
Aliso Canyon	Historically present	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
French Canyon	Historically present	Detected (14 April 2015)	12	8 (66.7%)	9 (75%)	3 (25%)
Cockleburr Canyon	Recently collected	Detected (13 April 2015)	4	1 (25%)	2 (50%)	2 (50%)
Upper Santa Margarita River	Unknown	Detected (13 April 2015)	8	0 (0%)	7 (88%)	0 (0%)
Middle Santa Margarita River	Unknown	Not Detected (17 April 2015)	3	0 (0%)	0 (0%)	0 (0%)
Santa Margarita Estuary	Historically present	Not Detected (13 &14 April 2015)	26	0 (0%)	1 (3.8%)	0 (0%)

Appendix A3 Table 2: Detections of **Black Crappie** on and near Camp Pendleton, California

Locations from Northern to Southern	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
			AC12s	Vert16s	COI
San Mateo Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
San Onofre Creek	Not Detected (15 April 2015)	6	0 (0%)	0 (0%)	0 (0%)
Las Pulgas (Las Flores) Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Aliso Canyon	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
French Canyon	Not Detected (14 April 2015)	12	0 (0%)	0 (0%)	0 (0%)
Cockleburr Canyon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Upper Santa Margarita River	Not Detected (13 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Middle Santa Margarita River	Not Detected (17 April 2015)	3	0 (0%)	0 (0%)	0 (0%)
Santa Margarita Estuary	Not Detected (14 April 2015)	26	0 (0%)	0 (0%)	0 (0%)
Lake O'Neil	Detected (17 April 2015)	5	2 (40%)	4 (80%)	3 (60%)

Appendix A3 Table 3: Detections of **Largemouth bass** on and near Camp Pendleton, California

Locations from Northern to Southern	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
			AC12s	Vert16s	COI
San Mateo Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
San Onofre Creek	Not Detected (15 April 2015)	6	0 (0%)	0 (0%)	0 (0%)
Las Pulgas (Las Flores) Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Aliso Canyon	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
French Canyon	Not Detected (14 April 2015)	12	0 (0%)	0 (0%)	0 (0%)
Cockleburr Canyon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Upper Santa Margarita River	Detected (13 April 2015)	8	0 (0%)	0 (0%)	2 (25%)
Middle Santa Margarita River	Not Detected (17 April 2015)	3	0 (0%)	0 (0%)	0 (0%)
Santa Margarita Estuary	Not Detected (14 April 2015)	26	0 (0%)	0 (0%)	0 (0%)
Lake O'Neil	Detected (17 April 2015)	5	2 (40%)	4 (80%)	0 (0%)

Appendix A3 Table 4: Detections of **Redeye bass** on and near Camp Pendleton, California

Locations from Northern to Southern	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
			AC12s	Vert16s	COI
San Mateo Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
San Onofre Creek	Not Detected (15 April 2015)	6	0 (0%)	0 (0%)	0 (0%)
Las Pulgas (Las Flores) Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Aliso Canyon	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
French Canyon	Not Detected (14 April 2015)	12	0 (0%)	0 (0%)	0 (0%)
Cockleburr Canyon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Upper Santa Margarita River	Detected (13 April 2015)	8	0 (0%)	4 (100%)	3 (75%)
Middle Santa Margarita River	Not Detected (17 April 2015)	3	0 (0%)	0 (0%)	0 (0%)
Santa Margarita Estuary	Not Detected (14 April 2015)	26	0 (0%)	0 (0%)	0 (0%)
Lake O'Neil	Not Detected (17 April 2015)	5	0 (0%)	0 (0%)	0 (0%)

Appendix A3 Table 5: Detections of **Arroyo toad** on and near Camp Pendleton, California

Locations from Northern to Southern	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
			AC12s	Vert16s	COI
San Mateo Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
San Onofre Creek	Not Detected (15 April 2015)	6	0 (0%)	0 (0%)	0 (0%)
Las Pulgas (Las Flores) Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Aliso Canyon	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
French Canyon	Not Detected (14 April 2015)	12	0 (0%)	0 (0%)	0 (0%)
Cockleburr Canyon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Upper Santa Margarita River	Detected (13 April 2015)	8	0 (0%)	4 (100%)	3 (75%)
Middle Santa Margarita River	Not Detected (17 April 2015)	3	0 (0%)	0 (0%)	0 (0%)
Santa Margarita Estuary	Not Detected (14 April 2015)	26	1 (3.8%)	0 (0%)	0 (0%)
Lake O'Neil	Not Detected (17 April 2015)	5	0 (0%)	0 (0%)	0 (0%)

Appendix A3 Table 6: Detections of **Green sunfish** on and near Camp Pendleton, California

Locations from Northern to Southern	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
			AC12s	Vert16s	COI
San Mateo Creek	Detected (15 April 2015)	8	6 (75%)	7 (88%)	0 (0%)
San Onofre Creek	Detected (15 April 2015)	6	0 (0%)	2 (33%)	0 (0%)
Las Pulgas (Las Flores) Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Aliso Canyon	Not Detected (15 April 2015)	8	0 (0%)	1 (13%)	0 (0%)
French Canyon	Not Detected (14 April 2015)	12	0 (0%)	0 (0%)	0 (0%)
Cockleburr Canyon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Upper Santa Margarita River	Detected (13 April 2015)	8	3 (38%)	3 (38%)	0 (0%)
Middle Santa Margarita River	Not Detected (17 April 2015)	3	0 (0%)	0 (0%)	0 (0%)
Santa Margarita Estuary	Not Detected (14 April 2015)	26	0 (0%)	0 (0%)	0 (0%)
Lake O'Neil	Not Detected (17 April 2015)	5	0 (0%)	0 (0%)	0 (0%)

Appendix A3 Table 7: Detections of **Western mosquitofish** on and near Camp Pendleton, California

Locations from Northern to Southern	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
			AC12s	Vert16s	COI
San Mateo Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
San Onofre Creek	Detected (15 April 2015)	6	1 (17%)	2 (33%)	0 (0%)
Las Pulgas (Las Flores) Creek	Not Detected (15 April 2015)	8	0 (0%)	1 (13%)	0 (0%)
Hidden Lagoon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Aliso Canyon	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
French Canyon	Not Detected (14 April 2015)	12	0 (0%)	0 (0%)	0 (0%)
Cockleburr Canyon	Detected (13 April 2015)	4	3 (75%)	3 (75%)	2 (50%)
Upper Santa Margarita River	Not Detected (13 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Middle Santa Margarita River	Not Detected (17 April 2015)	3	0 (0%)	0 (0%)	0 (0%)
Santa Margarita Estuary	Not Detected (14 April 2015)	26	0 (0%)	0 (0%)	0 (0%)
Lake O'Neil	Not Detected (17 April 2015)	5	0 (0%)	0 (0%)	0 (0%)

The cooler blank from Cockleburr Canyon had substantial contamination

Appendix A3 Table 8: Detections of **Brown bullhead** on and near Camp Pendleton, California

Locations from Northern to Southern	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
			AC12s	Vert16s	COI
San Mateo Creek	Detected (15 April 2015)	8	0 (0%)	8 (100%)	0 (0%)
San Onofre Creek	Detected (15 April 2015)	6	0 (0%)	2 (33%)	0 (0%)
Las Pulgas (Las Flores) Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Aliso Canyon	Not Detected (15 April 2015)	8	0 (0%)	1 (13%)	0 (0%)
French Canyon	Not Detected (14 April 2015)	12	0 (0%)	0 (0%)	0 (0%)
Cockleburr Canyon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Upper Santa Margarita River	Detected (13 April 2015)	8	0 (0%)	5 (63%)	0 (0%)
Middle Santa Margarita River	Not Detected (17 April 2015)	3	0 (0%)	1 (33%)	0 (0%)
Santa Margarita Estuary	Not Detected (14 April 2015)	26	0 (0%)	0 (0%)	0 (0%)
Lake O'Neil	Not Detected (17 April 2015)	5	0 (0%)	0 (0%)	0 (0%)

Appendix A3 Table 9: Detections of **Black bullhead** on and near Camp Pendleton, California

Locations from Northern to Southern	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
			AC12s	Vert16s	COI
San Mateo Creek	Detected (15 April 2015)	8	0 (0%)	8 (100%)	6 (75%)
San Onofre Creek	Not Detected (15 April 2015)	6	0 (0%)	0 (0%)	0 (0%)
Las Pulgas (Las Flores) Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Aliso Canyon	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
French Canyon	Not Detected (14 April 2015)	12	0 (0%)	0 (0%)	0 (0%)
Cockleburr Canyon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Upper Santa Margarita River	Not Detected (13 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Middle Santa Margarita River	Not Detected (17 April 2015)	3	0 (0%)	0 (0%)	0 (0%)
Santa Margarita Estuary	Not Detected (14 April 2015)	26	0 (0%)	0 (0%)	0 (0%)
Lake O'Neil	Not Detected (17 April 2015)	5	0 (0%)	0 (0%)	0 (0%)

Appendix A3 Table 10: Detections of **Common carp** on and near Camp Pendleton, California

Locations from Northern to Southern	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
			AC12s	Vert16s	COI
San Mateo Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
San Onofre Creek	Not Detected (15 April 2015)	6	0 (0%)	0 (0%)	1 (17%)
Las Pulgas (Las Flores) Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Not Detected (13 April 2015)	4	0 (0%)	1 (25%)	0 (0%)
Aliso Canyon	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
French Canyon	Not Detected (14 April 2015)	12	0 (0%)	0 (0%)	0 (0%)
Cockleburr Canyon	Detected (13 April 2015)	4	0 (0%)	2 (50%)	0 (0%)
Upper Santa Margarita River	Detected (13 April 2015)	8	8 (100%)	8 (100%)	8 (100%)
Middle Santa Margarita River	Not Detected (17 April 2015)	3	0 (0%)	0 (0%)	0 (0%)
Santa Margarita Estuary	Not Detected (14 April 2015)	26	0 (0%)	0 (0%)	0 (0%)
Lake O'Neil	Detected (17 April 2015)	5	1 (20%)	5 (100%)	0 (0%)

*All samples upstream of Camp Pendleton

Appendix A3 Table 11: Detections of **American bullfrog** on and near Camp Pendleton, California

Locations from Northern to Southern	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
			AC12s	Vert16s	COI
San Mateo Creek	Detected (15 April 2015)	8	4 (50%)	7 (88%)	4 (50%)
San Onofre Creek	Detected (15 April 2015)	6	2 (33%)	4 (67%)	0 (0%)
Las Pulgas (Las Flores) Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Aliso Canyon	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
French Canyon	Not Detected (14 April 2015)	12	0 (0%)	0 (0%)	0 (0%)
Cockleburr Canyon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Upper Santa Margarita River	Detected (13 April 2015)	8	4 (50%)	6 (75%)	6 (75%)
Middle Santa Margarita River	Detected (17 April 2015)	3	1 (33%)	1 (33%)	1 (33%)
Santa Margarita Estuary	Not Detected (14 April 2015)	26	0 (0%)	0 (0%)	0 (0%)
Lake O'Neil	Detected (17 April 2015)	5	0 (0%)	1 (20%)	1 (20%)

Appendix A3 Table 12: Detections of **Steelhead** on and near Camp Pendleton, California

Locations from Northern to Southern	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
			AC12s	Vert16s	COI
San Mateo Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
San Onofre Creek	Not Detected (15 April 2015)	6	0 (0%)	0 (0%)	0 (0%)
Las Pulgas (Las Flores) Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Aliso Canyon	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
French Canyon	Not Detected (14 April 2015)	12	0 (0%)	0 (0%)	0 (0%)
Cockleburr Canyon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Upper Santa Margarita River	Not Detected (13 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Middle Santa Margarita River	Not Detected (17 April 2015)	3	0 (0%)	0 (0%)	0 (0%)
Santa Margarita Estuary	Not Detected (14 April 2015)	26	0 (0%)	0 (0%)	0 (0%)
Lake O'Neil	Not Detected (17 April 2015)	5	0 (0%)	0 (0%)	0 (0%)

Appendix A3 Table 13: Detections of **Arroyo chub** on and near Camp Pendleton, California

Locations from Northern to Southern	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
			AC12s	Vert16s	COI
San Mateo Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
San Onofre Creek	Not Detected (15 April 2015)	6	0 (0%)	0 (0%)	0 (0%)
Las Pulgas (Las Flores) Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Aliso Canyon	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
French Canyon	Not Detected (14 April 2015)	12	0 (0%)	0 (0%)	0 (0%)
Cockleburr Canyon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Upper Santa Margarita River	Not Detected (13 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Middle Santa Margarita River	Detected (17 April 2015)	3	1 (33%)	2 (66%)	1 (33%)
Santa Margarita Estuary	Not Detected (14 April 2015)	26	0 (0%)	0 (0%)	0 (0%)
Lake O'Neil	Not Detected (17 April 2015)	5	0 (0%)	0 (0%)	0 (0%)

Appendix A3 Table 14: Detections of **Yellowfin Goby** on and near Camp Pendleton, California

Locations from Northern to Southern	eDNA status (date sampled)	eDNA Sampling effort (n)	Number of positive detections (% positive)		
			AC12s	Vert16s	COI
San Mateo Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
San Onofre Creek	Not Detected (15 April 2015)	6	0 (0%)	0 (0%)	0 (0%)
Las Pulgas (Las Flores) Creek	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Hidden Lagoon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Aliso Canyon	Not Detected (15 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
French Canyon	Not Detected (14 April 2015)	12	0 (0%)	0 (0%)	0 (0%)
Cockleburr Canyon	Not Detected (13 April 2015)	4	0 (0%)	0 (0%)	0 (0%)
Upper Santa Margarita River	Not Detected (13 April 2015)	8	0 (0%)	0 (0%)	0 (0%)
Middle Santa Margarita River	Not Detected (17 April 2015)	3	0 (0%)	0 (0%)	0 (0%)
Santa Margarita Estuary	Detected (14 April 2015)	26	0 (0%)	8 (31%)	0 (0%)
Lake O'Neil	Not Detected (17 April 2015)	5	0 (0%)	0 (0%)	0 (0%)

Appendix B

B1: Articles in peer-reviewed journals

Bold in print and open access
* in press
** in preparation

1. 2014 Environmental conditions influence eDNA persistence in aquatic systems
2. 2014 The room temperature preservation of filtered environmental DNA samples and assimilation into a Phenol-Chloroform-Isoamyl alcohol DNA extraction
3. 2014 Particle size distribution and optimal capture of aqueous macrobial eDNA
4. 2014 Quantifying environmental DNA signals for aquatic invasive species across multiple detection platforms
5. 2014 Improved methods for capture, extraction, and quantitative assay of environmental DNA from Asian bigheaded carp (Hypophthalmichthys spp.)
6. 2015 Fish environmental DNA is more concentrated in aquatic sediments than surface water
7. 2015 Quantification of mesocosm fish and amphibian species diversity via environmental DNA metagenetics
8. 2015 Improving confidence in environmental DNA species detection
9. 2016 The influence of stream bottom substrate on the retention and transport of vertebrate environmental DNA
10. 2016 Modeling the transport of environmental DNA using continuous flow-through column experiments
11. 2016 Estimating species richness using environmental DNA
12. 2016 Critical considerations for detection of aquatic species using environmental DNA
13. 2017 Fish community assessment with eDNA metagenetics: effects of sampling design and bioinformatic filtering
14. *2017 Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA
15. **2017 Monitoring stream fish diversity with environmental DNA (Eagle Creek)
16. **2017 Assessing the presence, absence, and detection error of endangered Tidewater Goby (*Eucyclogobius kristinae*) using molecular surveillance
17. **2017 Comparison of qPCR with high-through-put sequencing for detection of aquatic communities with environmental DNA (Camp Pendleton)

B2: Technical reports
None

B3: Conference or symposium proceedings
None

B4: Conference or symposium abstracts
1. Society for Freshwater Sciences, Sacramento CA, 2016

Title: When should you use genetic detection of a single species versus environmental DNA metagenetics in management?

Authors: Kristy Deiner, Yiyuan Li, Michael Pfrender, David Lodge, Chris Jerde

Abstract: Managers of natural resources in many cases know the species list of interest (endangered, invasive, etc.) for which they need monitoring data. When presented with the two prevailing genetic tools for species detection, PCR detection of single species versus environmental DNA (eDNA) metagenetics, there is little known about how the two methods compare and what the trade-offs are relative to monitoring a list of known species. Additionally, what is unknown is how well eDNA metagenetics performs in biodiversity monitoring in order to learn about what we do not know such as new detections of an invasive species. In this study we perform both methods side by side for several aquatic habitats for both endangered species or populations (i.e., Southern steelhead, Tidewater goby, Arroyo chub, Arroyo toad) and common invasive species (i.e., Common carp, American bullfrog) on the military base of Camp Pendleton, CA to ascertain the trade-offs such as financial considerations, accuracy and reliability. We will present these data in the context of key challenges managers may face when planning to implement a genetic monitoring plan for such species. We additionally develop a framework for how genetic detection of species can be used to develop baseline inventories for future risk assessments and management.

2. National Military Fish and Wildlife Association Annual Meeting, Omaha, NE, 2015

Title: Development of an environmental metagenetics approach for monitoring aquatic biodiversity.

Authors: Evans, N. T., M. A. Renshaw, C. R. Turner, B. P. Olds, Y. Li, C. L. Jerde, G. A. Lamberti, M. E. Pfrender, and D. M. Lodge.

Abstract: The objective of our research is to develop a practical environmental DNA (eDNA) and metagenetics-based method to assist management agencies in monitoring the status and trends in species richness on DoD lands. Specifically, our goal is to develop a bioassessment approach that not only detecting the presence or absence of a suite of target species but also provides statistically rigorous estimates of species richness for areas that could potentially be difficult to sample using traditional capture-based sampling techniques. To date, we have tested our metagenetic approach in an artificial mesocosm environment and in a natural stream fish assemblage. We successfully detected all the species present in the mesocosm experiment. In the stream ecosystem we detected all the species known to be present from electrofishing samples conducted simultaneously with the eDNA samples. Moreover, we detected additional species that may have either been missed by the electrofishing samples or species that inhabit locations upstream of the electrofishing sampling sites. Currently, we are working to sequence and analysis eDNA samples collected from a 2.2-ha pond located within Ft. Custer (Michigan National Guard) to compare with capture-based fish and amphibian samples. Following completion of this comparative analysis with the traditionally surveyed assemblages, we will apply our metagenetic approach in natural ecosystems with unknown species richness. We have Ft. Custer and Camp Pendleton as our test

locations for this initial application of our metagenetic approach with planned sampling in 2015.

3. American Fisheries Society 144th Annual Meeting, Quebec City, Canada, 2014

Title: Efficacy of eDNA-based metagenetic approaches for *en mass* molecular identification of fish and amphibian species richness.

Authors: Evans, N. T., B. P. Olds, M. A. Renshaw, C. R. Turner, C. L. Jerde, A. R. Mahon, M. E. Pfrender, G. A. Lamberti, and D. M. Lodge. 2014.

Abstract: Freshwater fauna are particularly sensitive to environmental change and disturbance. Management agencies frequently use fish and amphibian biodiversity as indicators of ecosystem health and a means to prioritize and assess management strategies. Traditional aquatic bioassessment that relies on capture or observation of organisms via nets, traps, and electrofishing gear typically has low detection probabilities for rare species and can injure individuals of protected species. Our objective was to determine if environmental DNA (eDNA) sampling and metagenetic analysis can be used to more accurately measure aquatic richness in species assemblages with differing community structures. We manipulated the density and relative abundance of eight fish and one amphibian species in a replicated mesocosm experiment. Environmental DNA was captured via filtered water samples and mitochondrial gene fragments were sequenced to measure species richness of each mesocosm. Metagenetic analysis identified all nine species in all treatment replicates. We discuss the relationship between sequencing read abundance and species biomass, the specificity of the metagenetic approach, and the effect of utilizing multiple genetic markers on species detection. Our results illustrate the potential for application of eDNA-metagenetic approaches for improved estimation of species richness in natural environments.

4. Ecological Society of America Annual Meeting Sacramento, California, USA, 2014

Title: Estimating species richness using environmental DNA

Authors: Jerde, CL, BP Olds, NT Evans, Y Li, MA Renshaw, K Deiner, AJ Shogren, E Andruszkiewicz, AR Mahon, CR Turner, JL Tank, D Bolster, GA Lamberti, ME Pfrender, and DM Lodge

Abstract: *Background/Question/Methods.* Using traditional capture methods, such as electrofishing and nets, to directly measure aquatic species richness is difficult when species are rare, so scientists use species richness estimators to account for undetected species. Environmental DNA (eDNA) is proving to be a robust indicator of rare, aquatic species presence. The metagenetics approach, that is the evaluation of taxon richness through homologous genes, potentially allows a water sample to reveal an aquatic system's species richness without the effort of traditional capture methods. Two main obstacles exist for developing metagenetic approaches for estimating species richness

with eDNA: quantifying errors in species detection, and defining the area over which an estimate of species richness relates. Here we quantify detection errors using the metagenetic approach where presence of species is known from electrofishing and conduct eDNA release studies in flowing environments to measure the retention and residence time. *Results/Conclusions:* The metagenetic approach is capable of detecting all 12 species found during electrofishing of four 60m reaches, but also identified species not observed in the study reaches. Two possible explanations for genetic detection of species not recovered using traditional gears is that DNA is sourced from fishes outside of the study area or species were not captured using electrofishing. From four independent trials of eDNA release, we show that eDNA is sticky and is retained in the environment much longer than conservative tracers. As a consequence, it is likely that if the DNA of an organism is found, then it is nearby. However, the residence time of our eDNA releases show that it is also plausible that DNA from outside the study reaches contribute to species richness estimates. This study demonstrates how eDNA with a metagenetic approach can provide accurate estimates of aquatic species richness and identifies limits to the inferences that can be made.

B5: Text books or book chapters

None

Appendix C

C1: Training videos available from workshop

Video content will be available from the Tools and Resources for Environmental DNA website (https://labs.wsu.edu/edna/) developed by our colleagues at Washington State University.

In the interim, the following links are available for viewing until the finalized versions are on the Technology transfer website:
Intro of Speakers: https://www.youtube.com/watch?v=GffZKr7BG08
Intro to Methods: https://www.youtube.com/watch?v=eYKGdZJAuQc
Case Study #1: https://www.youtube.com/watch?v=z2PNHMFlPsE
Case Study #2: https://www.youtube.com/watch?v=xaf4-olp3-I
Field Methods: https://www.youtube.com/watch?v=82idPQp70SU
Field Collection Demonstration: https://www.youtube.com/watch?v=JDonJj56cA4
What Happens in Lab: https://www.youtube.com/watch?v=gI9ne4to5so
Data Interpretation: https://www.youtube.com/watch?v=5-yRc1c19DA
Lawler Pond Interpretation: https://www.youtube.com/watch?v=C1rGhGjYPqI

C2: Protocol: Metagenetic Study of environmental DNA by two-step PCR amplification

Purpose: Detection of multiple species from environmental DNA samples using Illumina sequencing of multiplexed gene-specific PCR amplicons

Institution:	University of Notre Dame
Organization:	Genomics and Bioinformatics Core Facility
Phone:	(574) 631-1902
Address:	019 Galvin Life Science Center, Notre Dame, Indiana 46556, USA
Description:	Multiplexed gene-specific primers target conserved DNA regions and facilitates the capture of DNA signature of multiple species from a single water sample. Illumina sequencing allows for the effective screening of otherwise undetectable DNA signatures.
Last Updated:	05//2017

Facility Director: *Michael E. Pfrender*
Genomics and Bioinformatics Core Facility
University of Notre Dame
109B Galvin Life Science Center
(574) 631-0591
michael.pfrender.1@nd.edu

Assistant Director:
Melissa Stephen
University of Notre Dame
019 Galvin Life Science Center
(574) 631-1902
stephens.49@nd.edu

Postdoctoral Assistant:
Kristy Deiner
University of Notre Dame
019 Galvin Life Science Center
(408) 412 1704
alpinedna@gmail.com

Research Technicians:
Jacqueline Ann Lopez & Mark Renshaw
University of Notre Dame
019 Galvin Life Science Center
(574) 631-1902
jlopez11@nd.edu; mrenshaw@hpu.edu

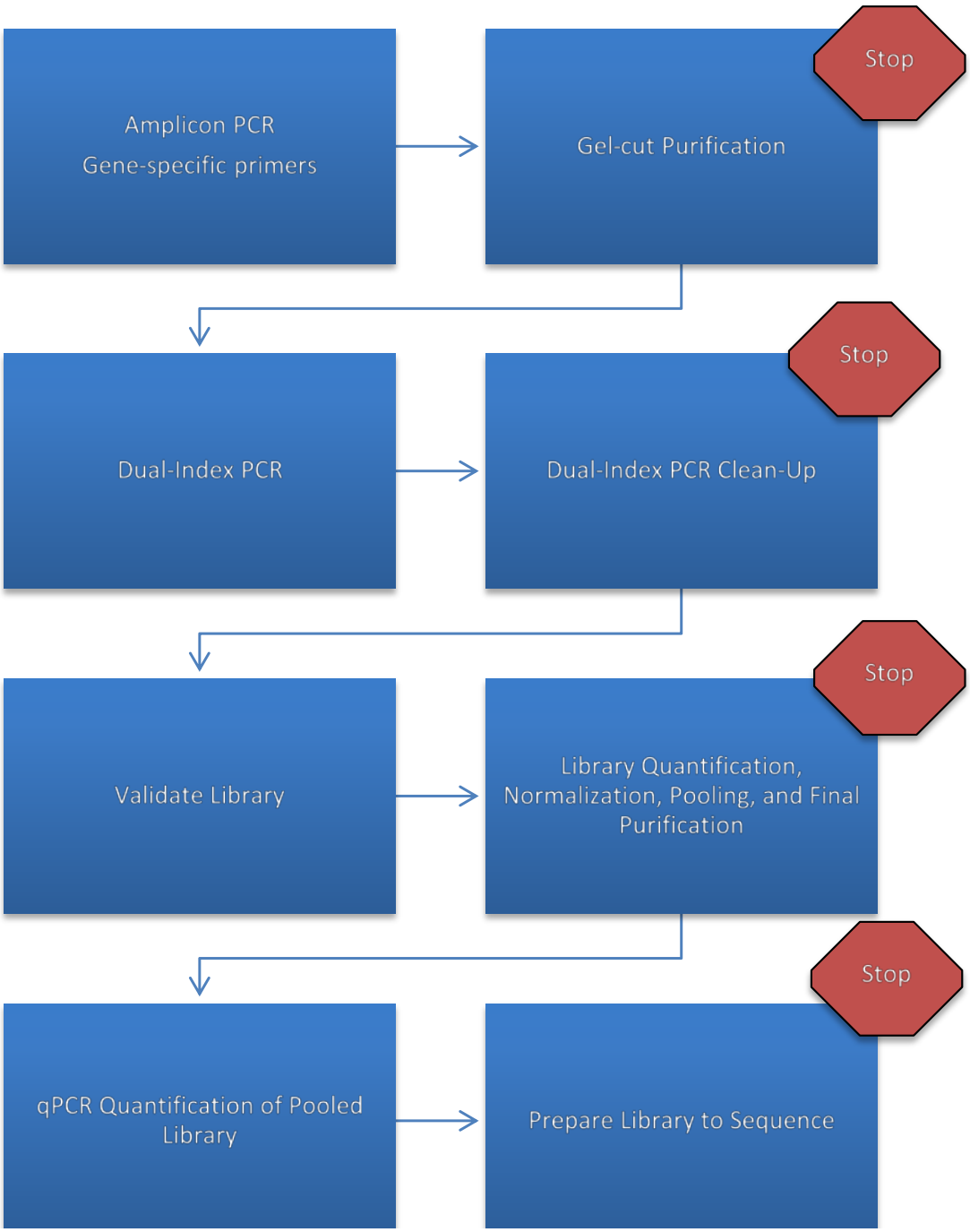
Table of Contents

Introduction

Primary Objectives

Develop a protocol for an Illumina-compatible DNA library of target-specific PCR amplicons derived from environmental water samples. Gene-specific primers are designed to target evolutionary conserved regions of the genome. Beginning with environmental water sample, target region is generated by PCR amplification. These DNA Amplicons are tagged with Illumina indexed adapters and sequenced in parallel with the Illumina MiSeq Instrument.

Figure 1 Library Preparation Workflow



Minimal Sample Information

- ***Sample ID –*** Unique, primary identifier of the environmental water sample
- ***Sample Location –*** GPS coordinates referencing the site from which the sample was taken.
- ***Sampling Date and time –*** Date and time when the sample was collected from the sample location.
- ***Sample Processing –*** Preparation of environmental water sample prior to DNA extraction (e.g., filtration, storage, etc.
- ***Sample DNA Extraction Method –*** Protocol detailing how DNA was isolated from the environmental water sample
- ***Amplicon PCR Method –*** Protocol to enrich target from environmental water sample (includes primer sequences, PCR conditions, and purification method
- ***Index PCR Method –*** Protocol to add index and Illumina flow cell sequence to Amplicon (includes primer sequences, PCR conditions, and purification method
- ***Sample Quality Assessment –*** Method to establish quantifiable metrics of quality for environmental water sample, amplicon product, indexed amplicon, and multiplexed library
- ***Modifications to Protocol by Sample ID –*** Report deviations from the predetermined/established protocol

Minimal Sample Control Information

- ***DNA Contamination Assessment Controls***
    - Sample Location Control Blank
    - Filtration control blank
    - Sample Extraction Control Blank
    - Amplicon PCR Control Blank
    - Index PCR Control Blank
- ***DNA Contamination Control by Sample ID Index***
    - Index file establishing relationship between Sample and DNA Contamination Assessment Controls (Sample Location Control Blank, Filtration control blank, Sample Extraction Control Blank, Amplicon PCR Control Blank, and Index PCR Control Blank)

Procedural Guidelines and Decontamination Procedure

It is very important to dress properly to maintain sterile and safe laboratory conditions. Foremost, avoid laboratory spaces contaminated with PCR products to prevent transfer contamination.

- To maintain a sterile and safe workspace, wear clean attire that is compliant with laboraroty standards and personal protective equipment (PPE) approved for the laboratory's biosafety level. In addtion, be aware of the these additional standards:
    - Do not wear any jewelry or clothing that may dangle;
    - Wash hands and face prior to entering the laboratory;
    - Tie back loose hair, use headcap;
    - Wear laboratory coat assigned to the preparation area, and secure the sleeves of the laboratory coat to prevent them from dragging across surfaces, as necessary;
    - Paints and close toed shoes must be warn and covered with shoe covers;
    - Change gloves often and pull over the sleeves of the laboratory coat.
- For each preparation area, dedicate specific:
    - Equipment and consumables,
    - Approved Personal Protective Equipment (PPE):
        - Laboratory coats,
        - Safety glasses,
        - Disposable face mouth masks,
        - Disposable powder-free nitrile gloves,
        - Face shield when interacting with UV-light sources.
- Decontaminate workspaces and equipment with freshly made 10% bleach solution before and after use.
    - Prepare fresh 10% bleach solutions prior to use because sodium hypochlorite, the active ingredient in bleach, breaks down when diluted.

    - Sodium hypochlorite in concentrated bleach is full-strength for up to 1-year from the date of manufacture when stored at room temperature (50˚C to 70˚C) and away from light. After 1-year of manufacture date, replace the bleach stock. Hospitals adhere to a 6-month post-production date replacement policy.
    - Manufature Date: On the Clorox® Regular-Bleach bottle, there is a two line ink-jet code with a 7-digit code on the topline and the EPA registration number on the second line. The first two digits are the plant identification code, followed by the 4-digit Julian date (last digit of the year and the Julian date of that year), and finally a two-digit shift identification.
        - A8809507 would be plant A8, year 8 or 2008, Julian day 095 or April 4th, and shift 07,
        - EPA registration number is 5813.
- After bleach wipe any surface that will come into contact with samples with 70% Ethanol. Must do this for metal surfaces as to not corrode the metal.
- Bleach and UV sterilize all workspaces and equipment for 30-minutes before and after use.
    - Keep a record of the UV light installation and replacement.
- Use reagent chillers that can be bleached and UV'd. Avoid use of wet ice when possible.
- Use only single-channel pipettors capable of disassembly and reassembly for cleaning, such as Eppendorf Research Plus Single-channel model. (Reference: Eppendorf Research Plus: Operating manual, 3120 900.012-06/042013. Do not use multi-channel pipettors at any point in the protocol.
- For all steps throughout the method, use only individually capped microcentrifuge tubes (0.5mL, 1.5mL, and 2.0mL) and PCR strips with individually attached caps. Unless specified, plate format (96-well or 384-well) is strongly discouraged.
- Replace gloves when transitioning
    - From one preparation area to another,
    - Between reagent primary stocks and working stocks,
    - Between working reagent stocks and DNA templates (i.e. extracted DNA and PCR products).
- Design a unidirectional workflow between preparation areas to avoid backtracking DNA between workspaces, especially those involving DNA manipulation.
- Establish a unidirectional workflow when handling PCR reagents and DNA templates. For example:
    - While wearing PPE and clean gloves, retrieve PCR reagents and Gene-Specific Primers from cold storage. Thaw in PCR designated reagent chiller. Assemble PCR master mix in a 1.5 mL tube. Dispense PCR master mix into reaction tube(s) and close reaction tube(s). Close PCR master mix tube and discard. Return PCR reagents and Gene-Specific Primers to cold storage. Discard gloves.
    - With clean gloves, retrieve DNA template(s) from cold storage. Thaw DNA template(s) in designated reagent chiller. Open PCR reaction tube, add DNA template, and then close both reaction tube and DNA tube. Repeat for all remaining PCR reactions. Return DNA template(s) to cold storage. Finally, discard gloves.
    - With clean gloves, transfer reagent chiller containing assembled PCR reactions to secondary containment and secure. Discard gloves. Transition to next preparation area.
    - With new gloves, decontaminate the exterior of the secondary container. Remove the reagent chiller containing the PCR reaction tube(s) and place on paper towels moistened with 10% Bleach Solution. Replace gloves and proceed with the workflow.

Amplicon PCR with Gene-Specific Primers

Gene-Specific Primers

Amplify target fragment(s) from eDNA sample with Illumina-compatible gene-specific primers. Illumina overhang adapter nucleotide sequence is added to the 5-prime end of each gene-specific primer.

Forward overhang: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[gene-specific sequence]
Reverse overhang: 5'- GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG -[gene-specific sequence]

Amplicon PCR Reaction Setup

- Setup work space and follow decontamination procedures
- Format: PCR strips with individually attached caps
- Batch size: Up to 64 reactions total (experimental samples plus controls)

Day Before PCR Setup:

1. Decontaminate surfaces by wiping with a paper towel moistened with 10% Bleach Solution from the spray bottle.
2. Decontaminate chiller racks, tube racks, and PCR racks by soaking in 10% Bleach Solution for 10 minutes. Rinse thoroughly with distilled water (DI) or reverse osmosis (RO) water. Allow to air dry overnight by laying flat and inverted ontop of and covered by fresh dry paper towels.

Day of PCR Setup:

1. UV sterilize both the bleach treated items and consumables for 30 minutes in a UV-equipped PCR hood.
2. After 30-minute UV sterilization is complete, only PCR reagents and DNA template will enter the PCR area. Do not introduce new eqipment or materials that have not be been bleach-treated and/or UV sterilized.

PCR Reaction Setup

1. Prepare the Master Mix and dispense into reaction tube(s).
    a. With clean gloves, retrieve PCR reagents (water, buffer, magnesium chloride, dNTP mix, and Primers) from cold storage. Thaw at room temperature. Vortex gently. Centrifuge briefly in mini centrifuge. Discard gloves.
    b. With clean gloves, assemble Master Mix in a 2.0 mL tube by adding the first 6 components as listed in Tables 3, 6, 9, 12, 15, and 18. Invert very gently to mix. Centrifuge briefly in the mini centrifuge. Set aside in chiller rack with PCR reagents.
    c. Retrieve Enzyme from cold storage. Centrifuge briefly. Add to Master Mix. Return the Enzyme to cold storage. Invert very gently to mix. Centrifuge briefly in the mini centrifuge. Set Master Mix aside in chiller rack with PCR reagents.
    d. Dispense Master Mix into first reaction tube and close its lid. Repeat until Master Mix is dispensed to all reaction tube(s). Close the lid of the Master Mix tube and discard. Keep the Nuclease-Free water for later use.
    e. Transfer the PCR reagents and Primers from the chiller rack to cold storage. Discard gloves.
2. Retrieve DNA template(s) from cold storage and add to PCR reaction tube(s).
    a. With clean gloves, retrieve DNA template(s) from cold storage. Thaw at room temperature. Vortex gently. Centrifuge briefly in mini centrifuge. Aspirate the volume of DNA template as listed in Tables 3, 6, 9, 12, 15, and 18 and then close its tube. Dispense the DNA template volume into the reaction tube and close its lid. Repeat until all DNA templates have been added to the appropriate reaction tube.
    b. For the PCR blanks, use the same nuclease-free water added to the Master Mix.
    c. Return the DNA template(s) back to cold storage. Discard gloves.
    d. With clean gloves, transfer the sample to secondary containment and transport assembled PCR reactions to the workspace containing the thermal cycler. Use the appropriate PCR program. Once the PCR program completes, proceed to next preparation area.

Table 1 L1492/H15149c Primer Loci Sequence and Fragment Size

Primer ID	Gene-Specific Sequence 5' – 3'	Loci / Target Region (bp)	Post-Amplicon PCR (bp) +67bp	Post-Index PCR (bp) +69bp
L14912	AAAAACCACCGTTGTTATTCAACTA	413	480	549
H15149c	GCCCCTCAGAATGATATTTGTCCTCA			

Table 2 L1492/H15149c Complete Primer Sequence with Illumina Overhang

Primer ID	Illumina-compatible gene-specific Primer Sequence 5' – 3'
L14912	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAAAACCACCGTTGTTATTCAACTA
H15149c	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCCCCTCAGAATGATATTTGTCCTCA

Table 3 L1492/H15149c Primer PCR Master Mix and PCR Cycling Conditions

L1492/H15149c Gene Reaction Conditions	
Component	**Amount**
Nuclease-free water	27.0 µl
HiFi Buffer, 5X, 7.5mM (BioRad)	10.0 µl
Magnesium chloride, 50mM (BioRad)	1.5 µl
dNTP mix, 10mM each	1.0 µl
Forward gene-specific primer, 10µM (IDT)	2.5 µl
Reverse gene-specific primer, 10µM (IDT)	2.5 µl
iProof HiFi DNA Polymerase, 2U/µl, (BioRad)	0.5 µl
DNA Template (concentration unknown)	5.0 µl
Total Volume	50.0 µl

L1492/H15149c Gene PCR Cycling Conditions			
Program: iProof25cycle		**Temperature (˚C)**	**Time (mm:ss)**
Step 1	Initial Denaturation	98.0	2:00
Step 2	Denature	98.0	00:10
Step 3	Anneal	60.0	00:20
Step 4	Extension	72.0	00:30
Step 5	Go to Step 2	9 cycles more	---
Step 6	Denature	98.0	00:10
Step 7	Anneal	58.0	00:20
Step 8	Extension	72.0	00:30
Step 9	Go to Step 6	9 cycles more	---
Step 10	Denature	98.0	00:10
Step 11	Anneal	55.0	00:20
Step 12	Extension	72.0	00:30
Step 13	Go to Step 10	29 cycles more	---
Step 14	Final Extension	72.0	10:00
Step 15	Hold	4.0	----

Table 4 Ac12S Primer Loci Sequence and Fragment Size

Primer ID	Gene-Specific Sequence 5' – 3'	Loci / Target Region (bp)	Post-Amplicon PCR (bp) +67bp	Post-Index PCR (bp) +69bp
Ac12S-F	ACTGGGATTAGATACCCCACTATG	385	452	521
Ac12S-R	GAGAGTGACGGGCGGTGT			

Table 5 Ac12S Complete Primer Sequence with Illumina Overhang

Primer ID	Illumina-compatible gene-specific Primer Sequence 5' – 3'
Ac12S-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTGGGATTAGATACCCCACTATG
Ac12S-R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAGAGTGACGGGCGGTGT

Table 6 Ac12S Primer PCR Master Mix and PCR Cycling Conditions

Ac12S Gene Reaction Conditions	
Component	**Amount**
Nuclease-free water	27.0 µl
HiFi Buffer, 5X, 7.5mM (BioRad)	10.0 µl
Magnesium chloride, 50mM (BioRad)	1.5 µl
dNTP mix, 10mM each	1.0 µl
Forward gene-specific primer, 10µM (IDT)	2.5 µl
Reverse gene-specific primer, 10µM (IDT)	2.5 µl
iProof HiFi DNA Polymerase, 2U/µl, (BioRad)	0.5 µl
DNA Template (concentration unknown)	5.0 µl
Total Volume	50.0 µl

Ac12S Gene PCR Cycling Conditions			
Program: iProof25cycle		**Temperature (˚C)**	**Time (mm:ss)**
Step 1	Initial Denaturation	98.0	2:00
Step 2	Denature	98.0	00:10
Step 3	Anneal	63.0	00:20
Step 4	Extension	72.0	00:30
Step 5	Go to Step 2	9 cycles more	---
Step 6	Denature	98.0	00:10
Step 7	Anneal	60.0	00:20
Step 8	Extension	72.0	00:30
Step 9	Go to Step 6	9 cycles more	---
Step 10	Denature	98.0	00:10
Step 11	Anneal	58.0	00:20
Step 12	Extension	72.0	00:30
Step 13	Go to Step 10	29 cycles more	---
Step 14	Final Extension	72.0	10:00
Step 15	Hold	4.0	----

Table 7 Am12S Primer Loci Sequence and Fragment Size

Primer ID	Gene-Specific Sequence 5' – 3'	Loci / Target Region (bp)	Post-Amplicon PCR (bp) +67bp	Post-Index PCR (bp) +69bp
Am12S-F	AGCCACCGCGGTTATACG	241	308	377
Am12S-R	CAAGTCCTTTGGGTTTTAAGC			

Table 8 Am12S Complete Primer Sequence with Illumina Overhang

Primer ID	Illumina-compatible gene-specific Primer Sequence 5' – 3'
Am12S-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCCACCGCGGTTATACG
Am12S-R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAAGTCCTTTGGGTTTTAAGC

Table 9 Am12S Primer PCR Master Mix and PCR Cycling Conditions

Am12S Gene Reaction Conditions	
Component	**Amount**
Nuclease-free water	27.0 µl
HiFi Buffer, 5X, 7.5mM (BioRad)	10.0 µl
Magnesium chloride, 50mM (BioRad)	1.5 µl
dNTP mix, 10mM each	1.0 µl
Forward gene-specific primer, 10µM (IDT)	2.5 µl
Reverse gene-specific primer, 10µM (IDT)	2.5 µl
iProof HiFi DNA Polymerase, 2U/µl, (BioRad)	0.5 µl
DNA Template (concentration unknown)	5.0 µl
Total Volume	50.0 µl

Am12S Gene PCR Cycling Conditions			
Program: iProof25cycle		**Temperature (˚C)**	**Time (mm:ss)**
Step 1	Initial Denaturation	98.0	2:00
Step 2	Denature	98.0	00:10
Step 3	Anneal	65.0	00:20
Step 4	Extension	72.0	00:30
Step 5	Go to Step 2	9 cycles more	---
Step 6	Denature	98.0	00:10
Step 7	Anneal	62.0	00:20
Step 8	Extension	72.0	00:30
Step 9	Go to Step 6	9 cycles more	---
Step 10	Denature	98.0	00:10
Step 11	Anneal	60.0	00:20
Step 12	Extension	72.0	00:30
Step 13	Go to Step 10	29 cycles more	---
Step 14	Final Extension	72.0	10:00
Step 15	Hold	4.0	----

Table 10 Ac16S Primer Loci Sequence and Fragment Size

Primer ID	Gene-Specific Sequence 5' – 3'	Loci / Target Region (bp)	Post-Amplicon PCR (bp) +67bp	Post-Index PCR (bp) +69bp
Ac16S-F	CCTTTTGCATCATGATTTAGC	330	397	466
Ac16S-R	CAGGTGGCTGCTTTTAGGC			

Table 11 Ac16S Complete Primer Sequence with Illumina Overhang

Primer ID	Illumina-compatible gene-specific Primer Sequence 5' – 3'
Ac16S-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTTTTGCATCATGATTTAGC
Ac16S-R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGGTGGCTGCTTTTAGGC

Table 12 Ac16S Primer PCR Master Mix and PCR Cycling Conditions

Ac16S Gene Reaction Conditions	
Component	**Amount**
Nuclease-free water	27.0 µl
HiFi Buffer, 5X, 7.5mM (BioRad)	10.0 µl
Magnesium chloride, 50mM (BioRad)	1.5 µl
dNTP mix, 10mM each	1.0 µl
Forward gene-specific primer, 10µM (IDT)	2.5 µl
Reverse gene-specific primer, 10µM (IDT)	2.5 µl
iProof HiFi DNA Polymerase, 2U/µl, (BioRad)	0.5 µl
DNA Template (concentration unknown)	5.0 µl
Total Volume	50.0 µl

Ac16S Gene PCR Cycling Conditions			
Program: iProof25cycle		**Temperature (˚C)**	**Time (mm:ss)**
Step 1	Initial Denaturation	98.0	2:00
Step 2	Denature	98.0	00:10
Step 3	Anneal	63.0	00:20
Step 4	Extension	72.0	00:30
Step 5	Go to Step 2	9 cycles more	---
Step 6	Denature	98.0	00:10
Step 7	Anneal	60.0	00:20
Step 8	Extension	72.0	00:30
Step 9	Go to Step 6	9 cycles more	---
Step 10	Denature	98.0	00:10
Step 11	Anneal	58.0	00:20
Step 12	Extension	72.0	00:30
Step 13	Go to Step 10	29 cycles more	---
Step 14	Final Extension	72.0	10:00
Step 15	Hold	4.0	----

Table 13 Ve16S Primer Loci Sequence and Fragment Size

Primer ID	Gene-Specific Sequence 5' – 3'	Loci / Target Region (bp)	Post-Amplicon PCR (bp) +67bp	Post-Index PCR (bp) +69bp
Ve16S-F	CGAGAAGACCCTATGGAGCTTA	310	377	466
Ve16S-R	AATCGTTGAACAAACGAACC			

Table 14 Ve16S Complete Primer Sequence with Illumina Overhang

Primer ID	Illumina-compatible gene-specific Primer Sequence 5' – 3'
Ve16S-F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGAGAAGACCCTATGGAGCTTA
Ve16S-R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAATCGTTGAACAAACGAACC

Table 15 Ve16S Primer PCR Master Mix and PCR Cycling Conditions

Ve16S Gene Reaction Conditions	
Component	**Amount**
Nuclease-free water	27.0 µl
HiFi Buffer, 5X, 7.5mM (BioRad)	10.0 µl
Magnesium chloride, 50mM (BioRad)	1.5 µl
dNTP mix, 10mM each	1.0 µl
Forward gene-specific primer, 10µM (IDT)	2.5 µl
Reverse gene-specific primer, 10µM (IDT)	2.5 µl
iProof HiFi DNA Polymerase, 2U/µl, (BioRad)	0.5 µl
DNA Template (concentration unknown)	5.0 µl
Total Volume	50.0 µl

Ve16S Gene PCR Cycling Conditions			
Program: iProof25cycle		**Temperature (˚C)**	**Time (mm:ss)**
Step 1	Initial Denaturation	98.0	2:00
Step 2	Denature	98.0	00:10
Step 3	Anneal	65.0	00:20
Step 4	Extension	72.0	00:30
Step 5	Go to Step 2	9 cycles more	---
Step 6	Denature	98.0	00:10
Step 7	Anneal	62.0	00:20
Step 8	Extension	72.0	00:30
Step 9	Go to Step 6	9 cycles more	---
Step 10	Denature	98.0	00:10
Step 11	Anneal	60.0	00:20
Step 12	Extension	72.0	00:30
Step 13	Go to Step 10	29 cycles more	---
Step 14	Final Extension	72.0	10:00
Step 15	Hold	4.0	----

Table 16 L2513/H2714 Primer Loci Sequence and Fragment Size

Primer ID	Gene-Specific Sequence 5' – 3'	Loci / Target Region (bp)	Post-Amplicon PCR (bp) +67bp	Post-Index PCR (bp) +69bp
L2513	GCCTGTTTACCAAAAACATCAC	202	269	338
H2714	CTCCATAGGGTCTTCTCGTCTT			

Table 17 L2513/H2714 Complete Primer Sequence with Illumina Overhang

Primer ID	Illumina-compatible gene-specific Primer Sequence 5' – 3'
L2513	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCCTGTTTACCAAAAACATCAC
H2714	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCCATAGGGTCTTCTCGTCTT

Table 18 L2513/H2714 Primer PCR Master Mix and PCR Cycling Conditions

L2513/H2714 Gene Reaction Conditions	
Component	**Amount**
Nuclease-free water	27.0 µl
HiFi Buffer, 5X, 7.5mM (BioRad)	10.0 µl
Magnesium chloride, 50mM (BioRad)	1.5 µl
dNTP mix, 10mM each	1.0 µl
Forward gene-specific primer, 10µM (IDT)	2.5 µl
Reverse gene-specific primer, 10µM (IDT)	2.5 µl
iProof HiFi DNA Polymerase, 2U/µl, (BioRad)	0.5 µl
DNA Template (concentration unknown)	5.0 µl
Total Volume	50.0 µl

L2513/H2714 Gene PCR Cycling Conditions			
Program: iProof25cycle		**Temperature (˚C)**	**Time (mm:ss)**
Step 1	Initial Denaturation	98.0	2:00
Step 2	Denature	98.0	00:10
Step 3	Anneal	60.0	00:20
Step 4	Extension	72.0	00:30
Step 5	Go to Step 2	9 cycles more	---
Step 6	Denature	98.0	00:10
Step 7	Anneal	58.0	00:20
Step 8	Extension	72.0	00:30
Step 9	Go to Step 6	9 cycles more	---
Step 10	Denature	98.0	00:10
Step 11	Anneal	55.0	00:20
Step 12	Extension	72.0	00:30
Step 13	Go to Step 10	29 cycles more	---
Step 14	Final Extension	72.0	10:00
Step 15	Hold	4.0	----

Amplicon PCR Reaction Gel-cut and Purification

Preparing the Agarose Gel

1) Assemble a horizontal electrophoresis chamber and a horizontal gel tray with combs.
    a. Electrophoresis Chamber: ThermoScientific Model: D3-14
        i. 30 (L) x 27 (W) x 11cm (H), 800mL Buffer Capacity
    b. Power Supply: Fisher Scientific, Model: FB300 0 – 100mA
    c. Gel Tray Type: 16 cm (L), 13 cm (W), UV transparent
    d. Comb Type: Tooth Count 24, Tooth Thickness 1.5mm, Volume metric: 25.0 µl
    e. Distance between rows: 4 cm
2) Prepare 200 mL of 0.5X TAE 2% w/v agarose gel in a 500 mL Erlenmeyer flask.
    a. Combine 4.0 gr of Agarose (OmiPure, PN EM2125) with 200 mL of 0.5X TAE Buffer in the 500 mL Erlenmeyer flask. Heat in the mixture in the microwave for 3 minutes [100% Power / High]. Swirl to mix. If agarose particles are still visible, continue to heat at 10-second intervals until the agarose has completely dissolved.
    b. Note: Continuously monitor as the agarose solution can boil over.
3) Allow agarose solution to cool to 60˚C. This will take approximately 5 minutes at ambient temperature.
    a. The agarose solution may have lost some water. Add distilled water to make up for lost volume.
4) Add the gel stain, swirl to mix, and pour the gel mixture into the gel tray to cast.
    a. Add 20.0 µl of SYBR Safe Gel Stain (Invitrogen, PN S33102) and swirl until thoroughly mixed.
    b. Pour the gel slowly into the cast. Displace bubbles with a pipette tip.
    c. Position four (4) combs at 4cm distance apart, and confirm the combs are seated correctly.
    d. Tent gel with aluminum foil to protect the SYBR Safe stain, as it is a light sensitive dye.
    e. Leave the agarose gel at room temperature for at least 1-hour to set completely.
5) Fill the horizontal electrophoresis chamber with 800 mL of 0.5X TAE Running Buffer.
6) Submerge the casted gel into the electrophoresis chamber. The gel should be at least 2 mm below the surface of the Running Buffer; else, add additional Running Buffer.
7) Remove the combs from the gel.

Preparing the Amplicon PCR product for loading into the agarose gel

1) Retrieve the DNA Ladder and Sample Loading Dye from cold storage.
    a. Ready-to-Use 100bp DNA Ladder, Biotium PN 31032A
    b. 6X Loading Dye, Biotium PN 99962
2) To each 0.2mL PCR reaction tube containing the 50µl of amplified PCR products, add 10.0 µl of 6X Loading Dye for 1X Loading Dye final concentration. Pipette up and down 10 times to mix thoroughly.
3) Load the DNA Standard and PCR products into the appropriate wells of the submerged gel.
    a. Well Position 1: Load 10.0 µl of Ready-to-Use 100bp DNA Ladder
    b. Well Position 2: Skip, leave empty
    c. Well Position 3: Load 27.0 µl of the first PCR product
    d. Well Position 4: Load remaining 27.0 µl of the first PCR product.
    e. Well Position 5: Skip, leave empty
    f. Well Position 6: Load 27.0 µl of the second PCR product
    g. Well Position 7: Load remaining 27.0 µl of the second PCR product.
    h. Well Position 8: Skip, leave empty
    i. Continue to load the samples in this manner, leaving an empty well in-between.
    j. Well Position 24: Load 10.0 µl of Ready-to-Use 100bp DNA Ladder
    k. Repeat this pattern for Row 2, 3, and 4.
4) Run the gel to separate the fragments.
    a. Secure the chamber with its lid.
    b. Connect the positive and negative leads to the power supply.
    c. Set the power supply to 100mV and timer for 60 minutes.
    d. Turn on the power, and check for bubbles to verify that the current is running.
5) After 60 minutes, turn off the power supply and unplug the positive and negative leads.
6) Remove the gel from the chamber and transfer to Alpha Innotech SA-1000 (red). Adjust the both UV setting and Zoom Out setting to the maximum value to record the banding pattern before excision of the PCR products in a single clear image.
7) Once the image is recorded, turn off the UV light, return to gel tray, and proceed to Amplicon excision.

Excision of Amplicon from the agarose gel

1) Setup a neon orange tube rack with labeled 2.0 mL centrifuge tubes. Open each collection tube and seat one gel excision tip in the mouth of the tube.
    a. Note: Neon orange tube rack is preferred because it glows in the dark when exposed to blue light, providing the user with the necessary light to handle the labeled tubes in the rack.
2) Transfer the gel to Dark Reader Transilluminator (Clarechemical, PN DR-46B).
3) Put on the orange-tinted glasses. Turn off all light sources and turn on the Dark Reader light source. PCR products should be visible on the gel in the form of glowing bands.
4) Excise the band from the gel. Use a new excision tip for each PCR product in the gel.
    a. With a gel excision tip seated on a P1000 set to 1000 µl, position the tip over the DNA band. Gently push the tip down into the gel to excise the band. To dispense the gel slice from the tip into the collection tube, push the pipette plunger down to the first stop. The gel slice should be expelled into the tube.
    b. Use the same excision tip and collection tube for additional gel slices of the same PCR product.
    c. NOTE: Take care when ejecting the gel slice as it may ricochet out of the tube or may become lodged inside of the tip. If the gel slice becomes lodged inside the tip, rapidly press the plunger to build up pressure, and the gel slice should be expelled from the excision tip.
5) Once all bands have been excised from the gel, record a final image to document the excision pattern.
    a. Transfer gel to Alpha Innotech SA-1000 (red). Adjust the both UV setting and Zoom Out setting to the maximum value to record the excision pattern in a single clear image. Dispose of the gel.
6) Record the weight of the gel slice(s) for each PCR product. Proceed with gel purification.
7) Purify the PCR products from the agarose using Qiagen QIAquick Gel Extraction Kit.
    a. Add 3 volumes of QG Buffer to 1 gel volume.
        i. 1 mL of liquid volume equals 1 gram of gel weight
        ii. Ex: 450 µl of QG Buffer per 150 mg of agarose gel.
    b. Incubate at 50˚C for 10 minutes. Vortex for 2 to 3 seconds every 3 minutes.
    c. After the gel slice has completely dissolved, the color of the solution will be yellow. If orange, add 10.0 µl of 3M Sodium acetate, pH 5, and mix. The mixture will turn yellow.
    d. Add 1 volume of Isopropanol. Invert to mix.
    e. Transfer up to 700 µl of the sample to the QiaQuick column with 2.0mL collection. Centrifuge for 1 minute at 13,000 x g. Discard flow-through. Repeat until all of the liquid has passed through the column.
    f. Add 750 µl of Wash Buffer to the QiaQuick column. Incubate for 5 minutes at room temperate to remove excess salts.
    g. Centrifuge for 1 minute at 13,000 x g. Discard flow-through.
    h. Centrifuge for 1 minute at 17,900 x g to remove residual ethanol from the filter.
    i. Transfer QiaQuick column to 1.5mL collection tube.
    j. To elute DNA, add 30.0 µl of Buffer EB to the filter. Incubate for 1 minute at room temperature. Centrifuge for 1 minute at 13,000 x g.
8) Measure sample concentration with Qubit HS DNA Assay kit.

Quantification of PCR products with Qubit

1. Prepare the Qubit Buffer and Qubit Reagent mixture according to manufacturer's guidelines.
2. Measure new standards for every batch of Qubit Buffer + Reagent mixture prepared.
3. For accurate and reproducible results, it is critical to homogenize the sample prior to quantification.
    a. Gently vortex the PCR product, and briefly centrifuge.
4. Assay 2.0 µl of PCR product to determine the concentration of the stock. Record the calculated stock concentration in ng/µl.

Dual-Index PCR with Nextera XT Index Primers

The PCR step adds Index 1 (i7) and Index 2 (i5) and sequences required for hybridization to the Illumina flow-cell and cluster generation.

Nextra XT Index (i7) Primer: 5'- CAAGCAGAAGACGGCATACGAGAT [i7]GTCTCGTGGGCTCGG
Nextera XT Index (i5) Primer: 5'- AATGATACGGCGACCACCGAGATCTACAC [i5] TCGTCGGCAGCGTC

Figure 15 Dual Index Set and Barcode Sequences

Index 1 (i7)	Index 1 Read Sequence	Index 1 Oligo Sequence	Index 2 (i5)	Index 2 Read Sequence	Index 2 Oligo Sequence
N701	TAAGGCGA	TCGCCTTA	S502	CTCTCTAT	CTCTCTAT
N702	CGTACTAG	CTAGTACG	S503	TATCCTCT	TATCCTCT
N703	AGGCAGAA	TTCTGCCT	S504	AGAGTAGA	AGAGTAGA
N704	TCCTGAGC	GCTCAGGA	S505	GTAAGGAG	GTAAGGAG
N705	GGACTCCT	AGGAGTCC	S506	ACTGCATA	ACTGCATA
N706	TAGGCATG	CATGCCTA	S507	AAGGAGTA	AAGGAGTA
N707	CTCTCTAC	GTAGAGAG	S508	CTAAGCCT	CTAAGCCT
N708	CAGAGAGG	CCTCTCTG	S517	GCGTAAGA	GCGTAAGA
N709	GCTACGCT	AGCGTAGC			
N710	CGAGGCTG	CAGCCTCG			
N711	AAGAGGCA	TGCCTCTT			
N712	GTAGAGGA	TCCTCTAC			

Read Sequences is the sequence generated within the Illumina reads.
Oligo Sequence is the sequence within the oligonucleotide primer when it's ordered.

Dual-index Strategy

One i7 index is paired with a single i5 index. The 96 possible combinations are illustrated in the table. Reserve a unique set of dual-index combinations for "no template control" or "PCR Blank" library because both "no template control" and "PCR Blank" libraries will be spiked into all lanes of sequencing. The remaining index combinations are randomly assigned to libraries.

Table 4 Dual-Index Matrix

	N701	N702	N703	N704	N705	N706	N707	N708	N709	N710	N711	N712
S502	N701-S502	N702-S502	N703-S502	N704-S502	N705-S502	N706-S502	N707-S502	N708-S502	N709-S502	N710-S502	N711-S502	N712-S502
S503	N701-S503	N702-S503	N703-S503	N704-S503	N705-S503	N706-S503	N707-S503	N708-S503	N709-S503	N710-S503	N711-S503	N712-S503
S504	N701-S504	N702-S504	N703-S504	N704-S504	N705-S504	N706-S504	N707-S504	N708-S504	N709-S504	N710-S504	N711-S504	N712-S504
S505	N701-S505	N702-S505	N703-S505	N704-S505	N705-S505	N706-S505	N707-S505	N708-S505	N709-S505	N710-S505	N711-S505	N712-S505
S506	N701-S506	N702-S506	N703-S506	N704-S506	N705-S506	N706-S506	N707-S506	N708-S506	N709-S506	N710-S506	N711-S506	N712-S506
S507	N701-S507	N702-S507	N703-S507	N704-S507	N705-S507	N706-S507	N707-S507	N708-S507	N709-S507	N710-S507	N711-S507	N712-S507
S508	N701-S508	N702-S508	N703-S508	N704-S508	N705-S508	N706-S508	N707-S508	N708-S508	N709-S508	N710-S508	N711-S508	N712-S508
S517	N701-S517	N702-S517	N703-S517	N704-S517	N705-S517	N706-S517	N707-S517	N708-S517	N709-S517	N710-S517	N711-S517	N712-S517

Duel-Index pairs reserved for “no template control” and “PCR blank” libraries

N705-S502	N705-S503	N703-S504	N706-S505	N702-S506	N711-S507	N702-S517
N709-S502	N707-S503	N711-S504	N708-S505	N707-S506	N704-S508	N706-S517
N701-S503	N712-S503	N701-S505	N710-S505	N703-S507	N709-S508	N712-S517

Combine Amplicons for Index PCR

As PCR is anticipated to amplify smaller fragments more efficiently than larger fragments, products from the Amplicon PCR are combined with regards to their size and subsequent anticipated relative performance (compared to other fragments in the mix) in the Index PCR. Given the concentration (ng/µl) obtained after Amplicon PCR, combine the six markers such that the final DNA concentration of the mix is composed of 1 part L2513/H2714, 1.25 parts Am12S, 1.5 parts Ve16S, 1.6 parts Ac16S, 2 parts Ac12S, and 3 parts L14912/H15149c. For example, if the Amplicon PCR produced the same 10 ng/µl DNA concentration for each of the six markers, the mixed DNA template could be composed of 1µl (10ng) L2513/H2714 (final concentration 0.097ng/µl), 1.25µl (12.5ng) Am12S (final concentration 0.121ng/µl), 1.5µl (15ng) Ve16S (final concentration 0.145ng/µl), 1.6µl (16ng) Ac16S (final concentration 0.155ng/µl), 2µl (20ng) Ac12S (final concentration 0.193ng/µl), and 3µl (30ng) L14912/H15149c (final concentration 0.290ng/µl).

Dual-Index PCR Reaction Setup

- Setup work space and follow decontamination procedures
- Format: PCR strips with individually attached caps
- Batch size: Up to 64 reactions total (63 experimental samples plus one (1) PCR Blank)

Day Before PCR Setup:

1. Decontaminate surfaces by wiping with a paper towel moistened with 10% Bleach Solution from the spray bottle.
2. Decontaminate chiller racks, tube racks, and PCR racks by soaking in 10% Bleach Solution for 10 minutes. Rinse thoroughly with distilled water (DI) or reverse osmosis (RO) water. Allow to air dry overnight by laying flat and inverted ontop of and covered by fresh dry paper towels.

Day of PCR Setup:

1. UV sterilize both the bleach treated items and consumables for 30 minutes in a UV-equipped PCR hood.
2. After 30-minute UV sterilization is complete, only PCR reagents and DNA template will enter the PCR area. Do not introduce new eqipment or materials that have not be been bleach-treated and/or UV sterilized.

PCR Reaction Setup

1. Prepare the Master Mix and dispense into reaction tube(s).
    a. With clean gloves, retrieve PCR reagents (water, buffer, magnesium chloride, and dNTP mix) and Index Primers from cold storage. Thaw at room temperature. Vortex gently. Centrifuge briefly in mini centrifuge. Discard gloves.
    b. With clean gloves, assemble Master Mix in a 2.0 mL tube by adding the first 5 components as listed in Table 8. Invert very gently to mix. Centrifuge briefly in the mini centrifuge. Set aside in chiller rack with PCR reagents.
    c. Retrieve Enzyme from cold storage. Centrifuge briefly. Add to Master Mix. Return the Enzyme to cold storage. Invert very gently to mix. Centrifuge briefly in the mini centrifuge.
    d. Dispense Master Mix into first reaction tube and close its lid. Repeat until Master Mix is dispensed to all reaction tube(s). Close the lid of the Master Mix tube and discard. Keep the Nuclease-Free water for later use.
    e. For each reaction tube, add the assigned i7 and i5 index primer.
    f. Transfer the PCR reagents and Index Primers from the chiller rack to cold storage. Discard gloves.
2. Retrieve mixed DNA template(s) from cold storage and add to PCR reaction tube(s).
    a. With clean gloves, retrieve mixed DNA template(s) from cold storage. Thaw at room temperature. Vortex gently. Centrifuge briefly in mini centrifuge. Aspirate the desired volume of mixed DNA template as listed in Table 8 and then close its tube. Dispense the mixed DNA template volume into the appropriate reaction tube containing the dual-index primer pair assigned to the sample and close its lid. Repeat until all DNA templates are added to the appropriate reaction tube.
    b. For the PCR blank, use the same nuclease-free water added to the Master Mix.
    c. Return the DNA template(s) back to cold storage. Discard gloves.
    d. With clean gloves, transfer the index PCR reactions to secondary containment and transport to the workspace containing the thermal cycler. Use the appropriate PCR program. Once the PCR program completes, proceed to next preparation area

Table 5 Dual Index PCR

Index PCR Reaction Conditions	
Component	**Amount**

Index PCR Cycling Conditions		
Program: iProof	**Temperature (˚C)**	**Time (mm:ss)**

Nuclease-free water	22.0 µl
HiFi Buffer, 5X (BioRad)	10.0 µl
Magnesium chloride, 50mM (BioRad)	1.5 µl
dNTP mix, 10mM each, (G-Bioscience)	1.0 µl
iProof HiFi DNA Polymerase, 2U/µl, (BioRad)	0.5 µl
Index i7 primer, 10µM	5.0 µl
Index i5 primer, 10µM	5.0 µl
DNA template	5.0 µl
Total Volume	50.0 µl

Step 1	Denaturation	98.0	2:00
Step 2	Denaturation	98.0	00:10
Step 3	Annealing	60.0	00:20
Step 4	Extension	72.0	0:30
Step 5	Go to Step 2	7 cycles more	---
Step 6	Extension	72.0	10:00
Step 7	Hold	4.0	---

Dual-Index PCR Reaction Clean up with Ampure XP Beads

Reference: 16S Metagenomic Sequencing Library Preparation: Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System, Part # 15044223 Rev. B

Note: Procedure is defined for 96-well plate format and 384-well plate format. Although the Agencourt Guide recommends 70% Ethanol for all washes, Illumina recommends 80% Ethanol for all washes. Adaptation of the procedure to 1.5mL microcentrifuge format requires no deviations from the manufacturer's guidelines.

Important: Prepare 80% Ethanol immediately prior to use.

Ampure XP Bead Purification

Component	Amount per sample
Ampure XP Beads, Agencourt	40.0 µl
Ethanol, 80% Solution	400.0 µl
Nuclease-free water	52.5 µl

1. Equilibrate the Agencourt Ampure XP bottle to room temperature for 30 minutes prior to use.
2. Vortex the Agencourt Ampure XP bottle to thoroughly resuspend the magnetic beads.
3. Add 40.0 µl of Ampure XP Beads to 1.5 mL tube.
4. Transfer the Index PCR reaction (50.0 µl) to the 1.5 mL tube containing the Ampure XP beads.
5. Mix the beads and PCR reaction thoroughly by pipetting up and down 10 times. Incubate for 5 minutes at room temperature.
6. Place the 1.5 mL tube on the Magnet Rack. Wait 2 minutes for the beads to bind to the magnet or until the solution is clear.
7. Keep the tube on the magnet. Set P200 pipettor to 85.0 µl. Aspirate slowly the supernatant and discard. Leave 5.0 µl of supernatant in the tube, else beads carry over with the supernatant.
8. Keep the tube on the magnet. Dispense 200.0 µl of 80% Ethanol Solution. Incubate for 30 seconds at room temperature. Aspirate slowly all of the ethanol solution and discard. Beads tend not to carry over as easily with the Ethanol Solution, so there is no need to leave supernatant in the tube.
9. Repeat the wash step once more for a total of two washes.
10. Aspirate any residual from the tube with a P10 Pipettor set to 10.0 µl.
11. Keep the tube on the magnet. Allow the beads to air dry for 2 to 5 minutes or until all of the Ethanol Solution has evaporated.

12. Remove the tube from the magnet. Add 32.5 µl of Nuclease-free water and pipette up and down 10 times to elute the PCR products from the beads. If the beads appear grainy upon resuspension, continue to pipette and down until the solution is homogeneous. Incubate for a minimum of 2 minutes but no longer than 5 minutes at room temperature.
13. Place the tube on the magnet. Incubate for 2 minutes to separate beads from the eluted PCR products.
14. Transfer 30.0 µl of the eluate to a new 1.5 mL tube. Do not attempt to transfer more volume. It is critical to avoid the beads as they interfere with down-stream applications.

Quantification of PCR products with Qubit

1. Prepare the Qubit Buffer and Qubit Reagent mixture according to manufacturer's guidelines.
2. Measure new standards for every batch of Qubit Buffer + Reagent mixture prepared.
3. For accurate and reproducible results, it is critical to homogenize the sample prior to quantification.
    a. Gently vortex the PCR product, and briefly centrifuge.
4. Assay 2.0 µl of PCR product to determine the concentration of the stock. Record the calculated stock concentration in ng/µl.

Validate Library

1. For eDNA samples, run 1.0 µl of the final library on Agilent Bioanalzyer DNA 7500 chip.
2. For No-template controls (NTC), run 1.0 µl of the final library on Agilent Bioanalyzer DNA High-Sensitivity Chip to evaluate the library.

Library Quantification, Normalization, and Pooling

Reference: 16S Metagenomic Sequencing Library Preparation: Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System, Part # 15044223 Rev. B

Calculate DNA concentration in nM, based on the qubit concentration and average size of DNA amplicons as determined by an Agilent DNA7500 Bioanalyzer trace from the previous step:

$$\frac{DNA\ Concentration\left(\frac{ng}{\mu l}\right)*10^6}{660\left(\frac{g}{mol}\right)*length\ (bp)} = \text{nM Concentration}$$

Dilute concentrated final eDNA library using nuclease-free water to 4 nM. eDNA Libraries were randomly assigned and evenly distributed by sampling location between two independent sequencing pools. Combine 5 µl of each 4nM diluted eDNA library for each mixed eDNA library. Negative controls are combined as a separate independent pool without normalization.

Remove primer contamination from the eDNA library and Negative control pools following Ampure Bead purification described in section "Dual-Index PCR Reaction Clean up with Ampure XP Beads" using 0.8X bead:sample ratio and elute in the same initial volume.

Determine the concentration of the purified eDNA library and negative control pools using Qubit dsDNA High-Sensitivity Assay. Assume the average size is unchanged. Adjust the pooled eDNA library concentration to 4 nM, if necessary. Negative control is unchanged.

Into each eDNA library pool, spike-in 12.5% v/v Negative control pool.

Perform final library quantitation using the KAPA Library Quantification Kit for Ilumina Platforms containing Kapa SYBR® FAST qPCR Master Mix. Dilute the eDNA library with spiked-in Negative Controls to 2pM in triplicate. Follow the KAPA Library Quantification Kit Illumina® platforms KR0405 – v6.14 Technical Data Sheet and Analysis Template to determine library concentration.

Prepare Library for Sequencing on MiSeq

Reference: 16S Metagenomic Sequencing Library Preparation: Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System, Part # 15044223 Rev. B.
Final molar concentration for each pool was based on values determined by the Qubit dsDNA, Bioanalyzer High-Sensitivity Chip Assay, and Kapa Illumina Library Quantification qPCR Assay. The final pool contained 87.5% eDNA Library Pool and 12.5% Negative Control Pool.

Denature 5.0ul of eDNA Library Pool with 5.0ul 0.2N NaOH. Mix. Incubate for 5 minutes at room temperature. Dilute with 990ul HT1 Buffer. Denatured eDNA Library Pool is now 20pM. Dilute 20pM Library solution with HT1 Buffer to either 7 pM with 20% v/v spike-in of 20pM PhiX Control v3 or 8pM with 10% v/v spike-in 20pM PhiX Control v3. Libraries are heat denatured prior to loading into reagent cartridge.

The final pool was sequenced on MiSeq Sequencer operating MiSeq Control Software v2.5 using MiSeq v3 600 cycle Kit. Sequencing format was 301 cycle read 1, 8 cycle index 1, 8 cycle index 2, and 301 cycle read 2. MiSeq Reporter v2.5 performed base calling by using Real Time Analysis (RTA) v1.18.54; the RTA output was demultiplexed according to the SampleSheet and converted to FastQ format with Illumina Bcl2fastq v2.18. Finally, BaseSpace Broker v2.1 reports the files to BaseSpace Sequencing Hub. Data for this project may be downloaded from BaseSpace Project Page.

References

1. Agencourt AMPure XP Information For Use Guide PCR Purification, PN B37419AA (August 2013)
2. 16S Metagenomic Sequencing Library Preparation: Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System, Part # 15044223 Rev. B
3. Qubit dsDNA HS Assay Kit, MAN0002326 | MP32851 Revision: B.0
4. QIAquick Gel Extraction Quick Start Protocol, July 2015
5. Evans NT, Olds BP, Renshaw MA, Turner CR, Li Y, Jerde CL, Mahon AR, Pfrender ME, Lamberti GA, Lodge DM (2015) Quantification of mesocosm fish and amphibian species diversity via environmental DNA metagenetics. Molecular Ecology Resources, doi:10.1111/1755-0998.12433.