

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

U.S. National Park Service Publications and
Papers

National Park Service

11-16-2018

Using redundant primer sets to detect multiple native Alaskan fish species from environmental DNA

Damian Menning

U. S. Geological Survey Alaska Science Center, dmenning@usgs.gov

Trey Simmons

National Park Service

Sandra Talbot

U. S. Geological Survey Alaska Science Center

Follow this and additional works at: <https://digitalcommons.unl.edu/natlpark>



Part of the [Environmental Education Commons](#), [Environmental Policy Commons](#), [Environmental Studies Commons](#), [Fire Science and Firefighting Commons](#), [Leisure Studies Commons](#), [Natural Resource Economics Commons](#), [Natural Resources Management and Policy Commons](#), [Nature and Society Relations Commons](#), [Other Environmental Sciences Commons](#), [Physical and Environmental Geography Commons](#), [Public Administration Commons](#), and the [Recreation, Parks and Tourism Administration Commons](#)

Menning, Damian; Simmons, Trey; and Talbot, Sandra, "Using redundant primer sets to detect multiple native Alaskan fish species from environmental DNA" (2018). *U.S. National Park Service Publications and Papers*. 206.

<https://digitalcommons.unl.edu/natlpark/206>

This Article is brought to you for free and open access by the National Park Service at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in U.S. National Park Service Publications and Papers by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



Using redundant primer sets to detect multiple native Alaskan fish species from environmental DNA

Damian Menning¹ · Trey Simmons² · Sandra Talbot¹

Received: 19 June 2018 / Accepted: 10 November 2018 / Published online: 16 November 2018

© This is a U.S. government work and its text is not subject to copyright protection in the United States; however, its text may be subject to foreign copyright protection 2018

Abstract

Accurate and timely data regarding freshwater fish communities is important for informed decision-making by local, state, tribal, and federal land and resource managers; however, conducting traditional gear-based fish surveys can be an expensive and time-consuming process, particularly in remote areas, like those that characterize much of Alaska. To help address this challenge, we developed and tested five multi-species environmental DNA (eDNA) primer sets for the simultaneous detection of up to 37 target fish species in a single sample. Using these primer sets can reduce the cost and time needed to perform future studies of fish communities. Our results comparing multiple samples from multiple lakes and streams using multiple next-generation sequencing runs show the efficacy and reproducibility of these primers.

Keywords eDNA · Metabarcoding · BLAST+ · Next generation sequencing · Illumina MiSeq

Introduction

In the vast, roadless landscapes that characterize much of Alaska, the use of traditional techniques to conduct comprehensive fish sampling in lakes and streams remains cost-prohibitive. As a result, data on the composition of fish communities in remote areas of Alaska tend to be scant. Nevertheless, such data are critical for the management of fisheries resources by local, state, tribal, and federal agencies. Recent developments in the identification of species using environmental DNA (eDNA) recovered from environmental samples (Valentini et al. 2009a, b) offer a powerful approach for fish community monitoring (McKelvey et al. 2016), or for any scientific study that relies on data regarding fish species richness or distribution (for a review see Thomsen and Willerslev 2015). Successful refinement and

application of eDNA techniques can provide robust, large-scale status and trend assessments of fish communities at a fraction of the cost of traditional fisheries surveys (Evans and Lamberti 2018). These kinds of data can also be useful in monitoring changes in the composition of fish communities as a result of changing Arctic and subarctic environments, and in elucidating how changes in fish communities may affect aquatic ecosystem health and piscivorous wildlife (Haynes et al. 2015; Laske et al. 2016).

Identification of species using eDNA is now an established method (Ficetola et al. 2008; Valentini et al. 2009a, b; Goldberg et al. 2011; Jerde et al. 2011; Andersen et al. 2012; Thomsen and Willerslev 2015), although early applications focused on the identification of single species, such as invasive or rare species, and employed traditional Sanger sequencing of barcoding genes (for example, the mitochondrial DNA (mtDNA) cytochrome oxidase I, COI) for fishes and amphibians (Hubert et al. 2008; Vences and Kohler 2008). More recently, much eDNA research has leveraged quantitative Polymerase Chain Reaction (qPCR) technologies (Spear et al. 2015; Laramie et al. 2015). While classical Sanger and qPCR methods have proved useful for identifying one to several target species (McKelvey et al. 2016; Spear et al. 2015), they are not feasible for enumerating biodiversity in bulk environmental samples which can contain DNA representing hundreds of species ranging from

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12686-018-1071-7>) contains supplementary material, which is available to authorized users.

✉ Damian Menning
dmenning@usgs.gov

¹ U. S. Geological Survey Alaska Science Center, 4210 University Drive, Anchorage, AK 99508, USA

² National Park Service, 4175 Geist Road, Fairbanks, AK 99709, USA

bacteria to vertebrates (Hajibabaei et al. 2011). An alternative approach to describing biodiversity in bulk environmental samples is the use of next-generation sequencing (NGS) methodologies in eDNA barcoding (metabarcoding; Taberlet et al. 2012) which, unlike traditional barcoding techniques, such as qPCR, can efficiently and cost-effectively separate individuals/species in support of biomonitoring programs that aim to inventory species assemblages from multiple temporally and spatially distinct environmental samples. As a result, metabarcoding as an ecological tool is burgeoning, and this approach is now beginning to be applied in the context of assessments of fish community diversity (Olds et al. 2016; Yamamoto et al. 2017; Evans et al. 2017; Evans and Lamberti 2018), and should facilitate biodiversity assessments at the landscape scale (Ficetola et al. 2008; Dejean et al. 2011; Goldberg et al. 2011; Jerde et al. 2011; Andersen et al. 2012).

We developed a NGS eDNA metabarcoding approach to expedite landscape-scale description of fish assemblages in streams and lakes from remote areas of Alaska. Here we present the results of the pilot phases of that research, which included (1) the selection, application, and testing of appropriate suites of barcoding indices based on downloaded sequence data; (2) the development of an Alaska specific reference database for the selected loci; (3) eDNA metabarcoding of a known community assemblage, sampled from an aquarium that contained a suite of fish species native to Alaskan waters; and (4) metabarcoding of unknown bulk samples collected from six lakes on the Arctic coastal plain of Alaska and from one lake and three streams in interior Alaska national parks (Table 1) at which traditional fisheries surveys were also conducted to provide a level of field validation. As part of this pilot research, to reduce time and project costs, and provide tools for researchers to design similar approaches elsewhere (Greiman et al. 2018), we developed a suite of simple user-friendly Python/Biopython scripts that can be used to design eDNA primers that target single or multiple species, to develop custom databases to reduce analysis time and data storage, and to compare unknown sequences to custom databases.

Materials and methods

Primer and reference database design

Primer design and reference database development were accomplished using Python (van Rossum 1995) and Biopython (Cock et al. 2009) scripts that are part of the U.S. Geological Survey Alaska Science Center Bioinformatics pipeline (Menning and Talbot 2018). The only a priori knowledge for the design of the primers was a list of freshwater fish species native to mainland Alaska and

a list of potential loci to be assayed (Supplemental Information 1). For this purpose, mainland Alaska was defined as the portion of the state north and west of Glacier Bay. Native species were defined as those freshwater fish species whose documented range included mainland Alaska, either in whole or in part. The primary reference for determining species range was Mecklenburg et al. (2002). Note that although the Alaska whitefish (*Coregonus nelsonii*), the lake whitefish (*C. clupeaformis*), and the humpback whitefish (*C. pidschian*) are listed as separate species in both Mecklenburg et al. (2002) and in the NCBI GenBank nucleotide repository, they are currently considered to belong to a single species complex (McDermid et al. 2007). Following Brown et al. (2012), we refer to these taxa collectively as *C. pidschian*. Note also that the presence of the pygmy whitefish (*Prosopium coulterii*) in Alaska is not documented in the formal literature, although Mecklenburg et al. (2002) list it as occurring.

The NCBI GenBank nucleotide repository (<http://www.ncbi.nlm.nih.gov>) was searched for each fish species for the following mitochondrial DNA (mtDNA) loci individually—12S, 16S, cytochrome b, and cytochrome oxidase—to identify which, if any, loci could be potential loci of interest for further study. The decision to focus on mtDNA loci was based on the greater prevalence of fish species mtDNA sequence information relative to nuclear DNA in GenBank, coupled with the finding that vertebrate mtDNA is more easily recovered than nuclear DNA in degraded samples (O'Rourke et al. 2000; Foran 2006). All FASTA sequences matching the fish taxa and the loci were downloaded from GenBank for primer design and potential inclusion in the reference database and pruned so that only unique taxon and sequence information were retained. If multiple identical sequences had different taxonomic information, those sequences were removed from the database unless one species was native to Alaska and the distributions of no others included mainland Alaska and were geographically discrete. Geographic discreteness was determined by comparing species range data obtained from FishBase.org. To meet the criterion of geographic discreteness, no overlap between the ranges of the species having identical sequence information was allowed. However, if multiple identical sequences had the same genus name but different species names, the sequence was kept in the database but was renamed *Genus* sp. (number). This increased the number of available sequences for those taxa.

The resulting condensed FASTA file was aligned using MEGA6 (Koichiro et al. 2013) and examined to identify conserved regions (potential primer sites), at least 17 nucleotides long, across taxa or taxonomic groups. Every possible potential primer site combination was tested against the aligned sequences to ensure primer specificity (matching our species of interest) and the ability to

Table 1 Sampling locations

General location	Name of water body	Type of water body	Traditional survey dates	eDNA survey dates	Number of eDNA samples	eDNA sample volume	Traditional survey reference
University of Alaska	UAF AQ	Aquarium	N/A	2013	1	250 mL	Andres Lopez, Curator of Fishes, University of Alaska Museum of the North, pers. comm. to TS
Arctic coastal plain	Crazy Bear Lake	Lake	2015	2015	2	2 L	Matthew Whitman, Bureau of Land Management, pers. comm. to TS
Arctic coastal plain	Hannah-Bear Lake	Lake	2015	2015	2	2 L	Matthew Whitman, Bureau of Land Management, pers. comm. to TS
Arctic coastal plain	INI-001	Lake	2014	2014	2	2 L	Matthew Whitman, Bureau of Land Management, pers. comm. to TS
Arctic coastal plain	INI-003	Lake	2014	2014	2	2 L	Matthew Whitman, Bureau of Land Management, pers. comm. to TS
Arctic coastal plain	INI-004	Lake	2014	2014	2	2 L	Matthew Whitman, Bureau of Land Management, pers. comm. to TS
Arctic coastal plain	INI-006	Lake	2015	2014	2	2 L	Matthew Whitman, Bureau of Land Management, pers. comm. to TS
Denali NP&P	Hogan Creek	Stream	1981, 2007, 2008	2013	1	2 L	Miller 1981, Simmons 2009, 2010
Denali NP&P	Igloo Creek	Stream	1981, 2007, 2008	2012, 2013, 2014	3	2 L	Miller 1981, Simmons 2009, 2010
Denali NP&P	Wonder Lake	Lake	2003, 2008	2013	1	20 L	Markis et al. 2004, Simmons 2010
Wrangell-St. Elias NP&P	Gilahina River	Stream	2003, 2008	2012, 2013, 2015	3	2 L	Markis et al. 2004, Simmons 2009

generate amplicons with enough sequence diversity that each sequence in the alignment can be assigned to a unique taxon. This process was conducted for all species on the species list, leading to the eventual selection of the mtDNA 12S [AK12S] and 16S [AK16S] loci as targets as well as the mtDNA COI for specific salmonid species (*Oncorhynchus* sp. [AKCOISal], *Coregonus* sp. [AKCOICor], and *Prosopium coulterii* [AKCOIPro] (Supplemental Information 1)). The resultant multi-species primers for use in PCR amplifications are shown in Table 2. We note that PCR reactions targeting *P. coulterii* used the salmonid COI forward primer (AKCOISalF) and a species specific reverse primer (AKCOIProR). All primers were checked for range of amplicon length (~ 150–250 base pairs were targeted) as well as

Table 2 List of primers designed for Alaskan fish species using Python/Biopython scripts

Locus	Sequence 5'–3'
AK16SF	CGAGAAGACCCTATGGAGC
AK16SR	GCGCTGTTATCCCTAGGGT
AK12SF	CTCGTGCCAGCCACCGCGGTTA
AK12SR	GGGTATCTAATCCCRGTTTG
AKCOISalF	TAGTATTTGGTGCCTGAGC
AKCOISalR	ATYATAACGAAGGCATGGGC
AKCOICorF	GCTGGTATTTCTCTATCTT
AKCOICorR	GCTGCTAGGACAGGAAGGGA
AKCOIProR	ATCATAACGAAGGCGTGGGC

primer–dimers, hairpins, and to ensure melting temperatures were within acceptable ranges (52–58 °C) using OligoAnalyzer 3.1 by Integrated DNA Technologies (<http://www.idtdn.com/calc/analyzer>). Newly designed primers were synthesized with Illumina adapter sequences by Eurofins Genomics (<http://www.eurofinsgenomics.com>) and validated with positive and negative PCR controls electrophoresed on 1% polyacrylamide gels. To multiplex 96 samples per MiSeq Illumina run, we combined Illumina sequence adapters, eight forward primer tags and 12 reverse primer tags, with our species/taxonomic group-specific primers. Based on the available sequence data, neither the three Alaskan *Cottus* species (*C. cognatus*, *C. aleuticus*, and *C. asper*) nor the three Alaskan *Salvelinus* species (*S. alpinus*, *S. malma*, and *S. namaycush*) can be reliably distinguished using these primers. Even though there are sequences available on GenBank for these species, there are not enough genetic differences in the available data to effectively differentiate between them. Consequently, we have designed *Cottus*-specific and *Salvelinus*-specific COI primers that allow separation of these species, but they have not yet been tested on field eDNA samples.

Once all of the 12S, 16S, and COI primers were verified, a reference database was developed by searching and downloading from GenBank all sequences that matched the search criteria of a ‘target locus’ and ‘eukaryotes’[porgn: __txid2759] or ‘bony fishes’[porgn: __txid7898]. These sequences were cropped based on the positions of the newly designed primers, leaving only the nucleotides including and between each of the primer pairs. The resulting data were examined to ensure that all of the primers aligned correctly with the target reference sequences and did not co-amplify non-target species. To avoid co-amplification of non-target species and resulting false positives, we used BLAST+ and the proposed primer sequences to search the NCBI database for non-target species that shared identical primer sequences with the target species. We retained in the reference database only sequences that contained both the forward and reverse primer sequences. These data were then examined for the presence of multiple identical sequences, as described above, and only one representative of each unique single sequence was retained. To further reduce the size of the reference database and reduce processing time we only included species of interest (i.e. species native to mainland Alaska) (Supplemental Information 2).

Sample collection, fishery surveys, DNA extraction, and Illumina sequencing

Water samples were collected in 2014 from six lakes on the Arctic coastal plain of Alaska, in 2012–2015 from two streams and another lake in Denali National Park and Preserve, and from a river in Wrangell-St. Elias National Park

and Preserve. Locations were selected where fish species data from traditional gear-based surveys were available (Table 1). In 2013, a water sample was collected from an aquarium located at the University of Alaska Fairbanks (UAF AQ); that sample represented a sample with known fish community composition. Environmental DNA samples were collected by filtering either 250 mL (aquarium) or 2 L (lakes and streams) of water through 0.45 µm cellulose acetate filter paper (Geotech) using a peristaltic pump in the field and then storing the filter paper and filtration residue in 5 mL Longmire Buffer (LMB) (Longmire et al. 1997) until processing. Species determinations of the captured fish from traditional fish surveys were conducted in the field.

Environmental samples were vortexed and eDNA was extracted using a 400 µL subsample of the LMB-preserved sample and a Qiagen DNeasy Blood and Tissue kit (Qiagen) following manufacturers suggested protocols, with the exception that volumes were doubled. To avoid contamination, all extractions were conducted in a laboratory in which Polymerase Chain Reactions (PCRs) have never been conducted and which is separated physically from laboratories in which PCRs are conducted. Each eDNA sample was subjected to PCRs in triplicate [8.6 µL dH₂O, 2.5 µL 10× PCR buffer, 2.0 µL MgCl (25 mM), 0.4 µL Taq Gold (5 µ/mL) (Life Technologies), 0.5 µL dNTP’s (10 mM) (ThermoFisher Scientific), 2.0 µL BSA (20 mg/ µL) (New England Biolabs), 2.5 µL of each primer (10 mM), and 4 µL sample eDNA]. PCR cycling parameters were as follows: 95 °C for 10 min, followed by 35 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 60 s, with a final elongation step at 72 °C continued for 30 min. Each locus was amplified separately. The three sets of PCR products from each locus were pooled by sample. Excess primers and dNTPs were removed from a 25 µL subaliquot of this pooled triplicate material using 2 µL of a 1:1 dilution of ExoSap (Affymetrix). The refined triplicate PCR products were quantified by fluorometry using a Quant-IT Broad Range kit (Thermo Fisher Scientific). Quantified triplicate PCR products were diluted to an equal concentration and 5 µL of each sample was pooled by locus. A 5 µL subaliquot of each of this pooled (by locus per individual) PCR product was electrophoresed on a 1% polyacrylamide gel to estimate PCR fragment sizes. A 30 µL subaliquot of the locus pooled PCR product was gel purified using a Pippin Prep (Sage Science), based on selecting previously-estimated target fragment sizes, so that primer dimers were avoided and as much product as possible was collected. These purified locus pooled PCR products were quantified by fluorometry using a Quant-IT High Sensitivity kit (Thermo Fisher Scientific) and diluted to 2 nM concentrations following Illumina guidelines (Illumina Document # 15039740 v01). Equal volumes of locus pooled PCR product were then pooled by sample and quantified by fluorometry using a Quant-IT High Sensitivity kit to verify dilution

concentration. This final sample/locus pooled product, which constituted the library to be subjected to next-generation sequencing, was diluted to 20 pM following the Illumina NextSeq Protocol A (Illumina Document #15048776 v02) for library dilution. All remaining steps followed the Illumina MiSeq protocol (Illumina Part #15034097 Rev. B). The eDNA library and PhiX were further diluted to 15 pM. Sequencing was performed using an Illumina MiSeq 300 cycle v2 reagent kit (2 × 151 paired-end cycle runs) (Illumina Part #MS-102-2002) on an Illumina MiSeq with a 30% PhiX spike.

Data analysis

All de-multiplexed data were retrieved from the Illumina MiSeq and analyzed independently using Python/Biopython scripts (Menning and Talbot 2018). This pipeline consisted of pairing forward and reverse reads using FLASH (<http://ccb.jhu.edu/software/FLASH/>, min-overlap=20, pthread-offset=33) and converting from FASTQ to FASTA. All paired reads in each de-multiplexed sample were filtered for size and primers, and analyzed independently. A local BLAST+ search using default parameters and an exact match comparison was conducted on all filtered/paired-end sequence data against the reference database (<https://www.ncbi.nlm.nih.gov/books/NBK279684/>). We used two methods to determine species identity: BLAST+ and exact match (100% identity). Because of lower stringency, BLAST+ searches can yield false positives, and because of within-taxon variability, exact match searches can yield false negatives. Therefore, we chose to use both approaches and compared results. Quality-filtering to remove sequencing errors was conducted by including only individual sequences that had a match

count greater than 0.01% of the total number of reads passing filter in that run (Bokulich et al. 2013).

Results

Primer and reference database design

All Illumina MiSeq data can be found at NCBI BioProject PRJNA389325. Sequences used to develop the databases were downloaded from NCBI GenBank on September 11th, 2017. Following the procedures above resulted in a reference database consisting of 157 unique sequences for the Alaska only database (Supplemental Information 2). This number (157) represent all sequences matching all the designed primers, and the search criteria “‘eukaryotes”[porgn: __txid2759]’ or “‘bony fishes”[porgn: __txid7898]’ and 12S, 16S, or COI and includes all fish species of interest. We note that some of the samples were held in LMB at room temperature for over 3 years which is significantly longer than the 150 days found by Wegleitner et al. (2015).

Aquarium community validation test

The UAF AQ sample was used as a positive control for all MiSeq runs (Table 3). Representatives of all six species detected in samples taken from UAF AQ were known to be present in the aquarium at the time of sampling with the exception of a single run that contained detections of rainbow trout (*Oncorhynchus mykiss*) and Arctic char (*Salvelinus alpinus*) (Andres Lopez, Curator of Fishes, University of Alaska Museum of the North, pers. comm. to TS). These detections of species not present at the time of sampling were only observed using the BLAST+ pipeline. As noted

Table 3 Data from three Illumina MiSeq runs using the University of Alaska aquarium sample (positive control)

Reference accession #	Locus hit	Genus species	MiSeq run	1		2		3	
				BLAST+	Exact	BLAST+	Exact	BLAST+	Exact
			Analysis method						
			Total reads passing filter	1,413,135		46,989		12,878	
			Common name						
AF333595.1	12S	<i>Catostomus catostomus</i>	Longnose sucker	15,908	996	2694	826	214	57
AB188191.1	12S	<i>Cottus aleuticus</i>	Coastrange sculpin	211 ^a	14 ^a				
NC_004592.1	12S	<i>Dallia pectoralis</i>	Alaska blackfish	6482	497	1960	631	275	76
KC844053.1	12S	<i>Lota lota</i>	Burbot	7859	929	1417	460	211	39
AF125509.1	12S	<i>Oncorhynchus mykiss</i>	Rainbow trout	1307					
AP013050.1	12S	<i>Prosopium cylindraceum</i>	Round whitefish	44,889	2955	8855	2544	1010	264
HQ167668.1	12S	<i>Salvelinus alpinus</i>	Arctic char	1729					
FJ872559.1	12S	<i>Thymallus arcticus</i>	Arctic grayling	60,385	5507	10,586	3432	1483	404
KJ866481.1	12S	<i>Thymallus arcticus</i>	Arctic grayling	775		41 ^a		28	

^aDid not meet the detection threshold of 0.01% of total sequences

Table 4 Data from multiple Illumina MiSeq runs from six Alaskan lakes on the Arctic Coastal Plain

Accession#	Locus hit	Genus species	Location	Crazy-Bear-2014-Run 1		Crazy-Bear-2014-Run 2		Hannah-Bear Lake-2014a-Run 1		Hannah-Bear Lake-2014a-Run 2		Hannah-Bear Lake-2014b-Run 1		Hannah-Bear Lake-2014b-Run 2		Lake INI-001-2014a-Run 1		Lake INI-001-2014b-Run 1	
				BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact
			Total reads passing filter	104,254	19,644	23,574	12,979	55,382	40,035	98,767	53,956								
			Common name																
AF333595.1	12S	<i>Catostomus commersoni</i>	Longnose sucker																
KT630723.1	COI-Cor	<i>Coregonus whitefish</i>	Broad whitefish	533	1534	2861	2647	75											
KT630719.1	COI-Cor	<i>Coregonus least</i>	Least Cisco	249	1018	476	793	1775											
AB188191.1	12S	<i>Cottus bairdii</i>	Coastal sculpin																
KJ778622.1	16S	<i>Cottus cognatus</i>	Slimy sculpin																
NC_004592.1	12S	<i>Dalia talaris</i>	Alaska blackfish	8 ^a	2 ^a	13 ^a	2 ^a												
NC_004593.1	12S	<i>Esox lucius</i>	Northern pike																
KC844053.1	12S	<i>Lota lota</i>	Burbot																
AF125509.1	12S	<i>Oncorhynchus mykiss</i>	Rainbow trout																
KP085590.1	12S	<i>Oncorhynchus mykiss</i>	Rainbow trout																
AP013050.1	12S	<i>Prosopium cylindraceum</i>	Round whitefish	2091	678	4900	3732	2281											
KJ128877.1	16S	<i>Pungitius pungitius</i>	Ninespine stickleback																
KJ627957.1	16S	<i>Pungitius pungitius</i>	Ninespine stickleback																

Table 4 (continued)

Accession#	Locus hit	Genus species	Location	Crazy-Bear-2014-Run 1		Crazy-Bear-2014-Run 2		Hannah-Bear-Lake-2014a-Run 1		Hannah-Bear-Lake-2014a-Run 2		Hannah-Bear-Lake-2014b-Run 1		Hannah-Bear-Lake-2014b-Run 2		Lake INI-001-2014a-Run 1		Lake INI-001-2014b-Run 1	
				BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact
			Total reads passing filter	104,254	19,644	23,574	12,979	55,382	40,035	98,767	53,956								
KJ627962.1	16S	<i>Pungitius pungitius</i>	Ninespine stickleback		27														
KJ627964.1	16S	<i>Pungitius pungitius</i>	Ninespine stickleback		66														
KR476959.1	16S	<i>Pungitius pungitius</i>	Ninespine stickleback		3071														
NC_011571.1	12S	<i>Pungitius pungitius</i>	Ninespine stickleback	77,279	20,681	2390	331	152	51	392	108	616	170	987	301	11,692	2158	2124	615
HQ167668.1	12S	<i>Salvelinus alpinus</i>	Arctic char			146	303					303		462					
FJ872559.1	12S	<i>Thymallus arcticus</i>	Arctic grayling			217	67					487		397	134				
			Total reads passing filter	61,241	23,280	54,025	145,294	23,384	53,398	95,717	55,202	74,163							
AF333595.1	12S	<i>Catostomus commersoni</i>	Longnose sucker				81												
KT630723.1	COI-Cor	<i>Coregonus nasus</i>	Broad whitefish	91										621		194			
KT630719.1	COI-Cor	<i>Coregonus nasus</i>	Least Cisco	3586	1407									1706	628				589

Table 4 (continued)

Accession#	Locus hit	Genus species	Location	Lake INI-001-2014a-Run 2		Lake INI-003-2014a-Run 1		Lake INI-003-2014b-Run 1		Lake INI-003-2014a-Run 2		Lake INI-003-2014b-Run 2		Lake INI-004-2014a		Lake INI-004-2014b		Lake INI-006-2014		
				BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+
			Analysis method																	
			Total reads passing filter	61,241	23,280	54,025	145,294	23,384	23,384	53,398	95,717	55,202	74,163							
AB188191.1	12S	<i>Cottus aleuticus</i>	Coastal-trange sculpin	678	198															
KJ778622.1	16S	<i>Cottus cognatus</i>	Slimy sculpin			8543	2804	24,967	7833	3280	1044	11,022	3352	284	84 ^a	27 ^a	7 ^a	1473	451	
NC_004592.1	12S	<i>Dalia pectoralis</i>	Alaska blackfish		68															
NC_004593.1	12S	<i>Esox lucius</i>	Northern pike											229	76 ^a					
KC844053.1	12S	<i>Lota lota</i>	Burbot																	
AF125509.1	12S	<i>Oncochilus mykiss</i>	Rainbow trout										68							
KP085590.1	12S	<i>Oncochilus mykiss</i>	Rainbow trout																	
AP013050.1	12S	<i>Prosopium cylindraceum</i>	Round whitefish	909	407															
KJ128877.1	16S	<i>Pungitius pungitius</i>	Ninespine stickleback																	
KJ627957.1	16S	<i>Pungitius pungitius</i>	Ninespine stickleback																	
KJ627962.1	16S	<i>Pungitius pungitius</i>	Ninespine stickleback																	
														3112	992	111	2187	258		

Table 4 (continued)

Accession#	Locus hit	Genus species	Location	Lake INI-001-2014a-Run 2		Lake INI-001-2014b-Run 2		Lake INI-003-2014a-Run 1		Lake INI-003-2014b-Run 1		Lake INI-003-2014a-Run 2		Lake INI-003-2014b-Run 2		Lake INI-004-2014a		Lake INI-004-2014b		Lake INI-006-2014		
				BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+
			Total reads passing filter	61,241	23,280	54,025	145,294	23,384	95,717	53,398	95,717	55,202	74,163									
KI627964.1	16S	<i>Pungitius pungitius</i>	Nines-pine stickleback																			
KR476959.1	16S	<i>Pungitius pungitius</i>	Nines-pine stickleback																			
NC_011571.1	12S	<i>Pungitius pungitius</i>	Nines-pine stickleback	4126	841	563	85	3336	1088	28,154	8059	1664	521	12,632	3771	3782	805	1328	354	3195	970	
HQ167668.1	12S	<i>Salvelinus alpinus</i>	Arctic char									332	189	299	2 ^a	1641						
FJ872559.1	12S	<i>Thymallus arcticus</i>	Arctic grayling			36 ^a	12 ^a															

Sample runs are labeled with the name of the site, followed by the year of eDNA sample collection. The collection dates are listed as 2014a and 2014b, corresponding to sample collection dates in June and July 2014, respectively. In cases where DNA extracted from the same sample was included in separate MiSeq runs, these are labeled as Run 1 and Run 2. Boldface type indicates sequence detections corresponding to species documented at that site in at least one traditional fisheries survey

^aDid not meet the detection threshold of 0.01% of total sequences

above, the three *Cottus* species that occur in Alaska (*Cottus cognatus*, *C. aleuticus*, and *C. asper*) cannot be reliably distinguished using either the 12S or 16S locus. Although *C. cognatus* was present in the aquarium, a 12S sequence from *C. aleuticus* was detected in only one of the three MiSeq runs. This suggests that while the 12S primer may not be able to differentiate between the three *Cottus* sp. found in Alaska, it can identify to the genus level and if site information is available, a tentative species determination may still be possible. Without prior site information, determination to genus would still be possible.

Field survey eDNA verification

Data from single Illumina MiSeq runs using a single sample (Lake INI-006, Hogan Creek), multiple MiSeq runs using a single sample (Crazy Bear Lake, Wonder Lake), single MiSeq runs using multiple samples (Igloo Creek, Gilahina River, Lake INI-004), and multiple MiSeq runs using multiple samples (Hannah-Bear Lake, Lake INI-001, Lake INI-003) are shown in Tables 4 and 5. Only individual sequences that had a match count greater than 0.01% of the total number of reads passing filter in that run are included. The numbers listed indicate the total number of BLAST+ or exact match hits of the sample sequences to the reference database.

We successfully recovered sufficient sequence data from 89% of the fish species that have been reported in these systems when the data were analyzed using BLAST+. In each instance in which either multiple runs or multiple samples were analyzed, the expected species were detected in more than one run or sample. The exceptions were the Gilahina River, where we failed to detect either *O. kisutch* or *O. tsawyscha* in any of three samples and Lake INI-004, where we failed to detect *Coregonus pidschian* or *Thymallus arcticus* in either of two samples. As noted above, the *Cottus aleuticus* 12S sequence detected at Hogan Creek, Wonder Lake, Igloo Creek, Hannah-Bear Lake, Lake INI-001, and the Gilahina River does not reliably indicate the presence of that species in those samples; based on the known species ranges (Mecklenburg et al. 2002) and the results from traditional fisheries surveys we are confident that these detections instead correspond to *C. cognatus*. Similarly, the *Salvelinus* sp. sequence detections in Wonder Lake, Hannah-Bear Lake, Lake INI-003, Lake INI-004, Lake INI-006 and the Gilahina River do not necessarily indicate the presence of the corresponding species. Based on habitat preferences, known species ranges, and the results of the traditional fisheries surveys, we are confident that the *Salvelinus* sp. 3 detected in Wonder Lake, Lake INI-004 and Lake INI-006 is *S. namaycush*, and in the Gilahina River is *S. malma*.

We also detected eDNA at most sites from additional species that were not observed during traditional fisheries surveys listed in Table 1. In almost all cases, these

detections were of species that, based upon habitat preferences and known species ranges, would reasonably be expected to occur in those systems. For the Arctic coastal plain lakes, *C. cognatus*, *Lota lota*, *C. nasus*, *Dalia pectoralis*, *S. namaycush*, *S. malma*, *S. alpinus*, *P. cylindraceum*, *Catostomus catostomus*, and *T. arcticus* have all been reported either in other nearby lakes from the current study (Matthew Whitman, Bureau of Land Management, pers. comm.) or from other studies conducted in the area (Mecklenburg et al. 2002; Haynes et al. 2014; Laske et al. 2016). Nearly all of these undocumented but reasonable species were detected in more than one sample or more than one MiSeq run from each site. Based on habitat preferences, the *Salvelinus* sp. 3 detected in Lake INI-003 could correspond to the presence of either *S. alpinus* or *S. namaycush*, or both. Because Hannah-Bear Lake is a flow-through lake where riverine eDNA might also be present, any of the three *Salvelinus* species could have been detected in samples from that site. In several other instances, unexpected detections were observed, including *C. clupeaformis* in Hogan Creek, *Esox lucius* and *Pungitius pungitius* in Igloo Creek, and *O. mykiss* in Lake INI-003. Based upon their habitat preferences and known species ranges, it is unlikely that eDNA from these species was actually present in these systems. In every case, these unexpected species were only detected in a single sample or MiSeq run.

Discussion

We used Python/Biopython scripts (Menning and Talbot 2018) to design novel primers that target multiple fish species, to create primer specific reference databases, and to match unknown sequences to these databases using NCBI BLAST+ or an exact match script on a local computer. The multi-species primer sets were used to multiplex loci in NGS metabarcoding analyses. We accurately and repeatedly detected all of the species present in the UAF aquarium (save *C. aleuticus*, although genus-level identification was possible), as well as the majority of species reported using traditional fisheries methods in six lakes on the Arctic Coastal Plain of Alaska as well as in one lake and three streams in interior Alaska national parks. The bioinformatics pipeline used to perform sequence identification incorporated a custom reference database that only included unique sequences corresponding to the species of interest, and also compared two different sequence similarity searching methods. The first method used the reference database to search for exact matches within the MiSeq data whereas the second method used BLAST+ to search the MiSeq data while allowing for some mismatches between sequences. Both methods have advantages and disadvantages. Generally, the BLAST+ search resulted in the detection of more species and more

Table 5 Data from multiple Illumina MiSeq runs from three Alaskan streams and one lake in Alaskan National Parks and Preserves

Accession#	Locus hit	Genus species	Loca- tion:	Hogan Creek- 2013		Igloo Creek- 2012		Igloo Creek- 2013		Igloo Creek-2014		Gilahina River- 2012		Gilahina River- 2013		Gilahina River- 2015		Wonder Lake- 2013-Run 1		Wonder Lake- 2013-Run 2	
				Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+
			Analysis method:	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact
			Total reads passing filter:	133,082	17,865	20,665	193,123	4922	101,220	37,763	104,970	13,553									
			<i>Com- mon name</i>																		
KR733409.1	COIPro	<i>Coregonus humpback whitefish</i>	158																		
AB188191.1	12S	<i>Cottus aleuticus</i>	17,457	1 ^a	4361	565	5120	808	114,253	15,112	57	11	10,768	1401	866	1 ^a	56,912	9400	46	1 ^a	
KJ778622.1	16S	<i>Cottus slimy</i>	8820																		89
NC_004593.1	12S	<i>Esox lucius</i>	30																		
KC844053.1	12S	<i>Lota lota</i>																			
AF125509.1	12S	<i>Oncorhynchus mykiss</i>																			
KP085590.1	12S	<i>Oncorhynchus mykiss</i>																			
NC_011571.1	12S	<i>Pungitius pungitius</i>																			
HQ167668.1	12S	<i>Salvelinus alpinus</i>																			
EU522418.1	COISal	<i>Salvelinus namaycush</i>																			
salvelinus3	COISal	<i>Salvelinus sp.3</i>																			
F1872559.1	12S	<i>Thymallus arcticus</i>	3773		3459	267	6725	480													1127

Table 5 (continued)

Accession#	Locus hit	Genus species	Location	Hogan Creek-2013		Igloo Creek-2012		Igloo Creek-2013		Igloo Creek-2014		Gilahina River-2012		Gilahina River-2013		Gilahina River-2015		Wonder Lake-2013-Run 1		Wonder Lake-2013-Run 2	
				BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact
			Analysis method:	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact	BLAST+	Exact
			Total reads passing filter:	133,082	17,865	20,665	193,123	4922	101,220	37,763	104,970	13,553									
KJ866481.1	12S	<i>Thymallus arcticus</i>	Arctic grayling	19	78																

Sample runs are labeled with the name of the site, followed by the year of eDNA sample collection. In cases where DNA extracted from the same sample was included in separate MiSeq runs, these are labeled as Run 1 and Run 2. Boldface type indicates sequence detections corresponding to species documented at that site in at least one traditional fisheries survey

^aDid not meet the detection threshold of 0.01% of total sequences

total hits than the exact match search. There were no misidentifications with the exact match search, whereas even with a filtered, custom reference database, the BLAST+ search was unable to differentiate closely related species such as *Cottus* sp. and *Salvelinus* sp. and there were several other obviously erroneous identifications (Tables 4, 5).

In a few cases, expected species were not detected. Specifically, in lake INI-004 *C. pidschian* and *T. arcticus* and in the Gilahina River *O. kisutch* and *O. tshawytscha* were not detected in our assay, although they had been documented in traditional fish surveys. There are several possible explanations for the failure to detect these species using eDNA. First, individuals of those species may not have been present upstream of the sampling site in the Gilahina River, or too far upstream to provide an adequate concentration of eDNA to allow detection using this assay (Tillotson et al. 2018). Similarly, in Lake INI-004, individuals of these species may not have been present in the lake at the time of eDNA sampling or may not have been located close enough to the eDNA sampling sites to provide adequate eDNA. This effect would be more pronounced for species that are present in low abundance (Tillotson et al. 2018). Collection of replicate samples from multiple locations at a site of interest (e.g., longitudinally in a river, or from different habitats in a lake) and at different times of the year is one way to minimize the potential for false negatives in eDNA assays. The detection of multiple species at other sites that, although reasonably expected to be present based on habitat preferences and species range, have not been reported from those sites using traditional survey methods, suggests that in some cases eDNA-based surveys may be more sensitive than traditional survey methods, as has been shown in other studies (Eiler et al. 2018; Hinlo et al. 2017; Wittwer et al. 2018).

In several cases, we detected unexpected species using BLAST+. Based on the habitat preferences and known ranges of these species, it is unlikely that they were present at the sites where they were detected using eDNA. We suggest that these detections are most likely due to misidentifications of the sequences by BLAST+ in the 12S or COIPro loci due to regional mtDNA sequence variability and the absence of those variants in the reference database. Alternatively, some of the sequences may have been incorrectly attributed in GenBank, as the presence of misidentified sequences in GenBank is a known issue (Nilsson et al. 2006; Shen et al. 2013). Collection of replicate samples, or utilizing multiple MiSeq runs if only a single sample is available, may minimize the likelihood of false positive species identifications. In every case in which unexpected species were detected, the mistaken detections occurred in only one sample (*O. mykiss* in Lake INI-003, *E. lucius* and *P. pungitius* in Igloo Creek), or in a single MiSeq run (*O. mykiss* and *S. alpinus* in the aquarium sample).

These misidentifications may be corrected by increasing the stringency of the BLAST+ search criteria, developing more specific species primers, requiring detections with multiple markers and/or in multiple sample replicates (Evans et al. 2017), or by increasing the diversity of sequences in the reference database. The lack of sequence diversity in the reference database was also problematic for the exact match search. As noted above the exact match did not misidentify any sequences, but in some cases did not detect species that were identified by BLAST+ and/or traditional gear-based fish surveys. This is because for some species only a single, or few, unique sequences were included in the reference database for any given marker, meaning that intraspecies sequence variability was poorly represented. For example, in the UAF Aquarium and Igloo Creek samples (Tables 3, 5), the exact match search found *Thymallus arcticus* (FJ872559.1) but missed *Thymallus arcticus* (KJ866481.1). The amplicons generated by the 12S primers are 200 bps long and there are two differences between these two sequences representing a 99% sequence similarity (Supplemental Information 3). These differences were found by BLAST+ but not by the exact match search criteria. Accordingly, we recommend that, at least during the initial phases of a project, tissue samples be collected from multiple specimens for species of interest at a subset of sites in the general area in which an eDNA survey is being conducted and sequence data from the target loci be obtained for those species (Rodgers et al. 2017).

The results of multiple runs of our positive control (UAF AQ; see Table 3) show that although there is variability in the total number of sequences and sequence matches, all expected species were consistently correctly identified. The only unexpected result from the UAF AQ sample analysis was the finding of sequences attributed to *S. alpinus* and *O. mykiss* in a single MiSeq run using BLAST+. The detection of these two species in a single MiSeq run may be due to inadvertent contamination between samples in the lab or in the field, although subsequent runs of the same sample did not show any indication of this possible cross-contamination. The most plausible explanation, due to the fact that these species were not detected using the exact match Python script, is misidentifications by BLAST+ of closely related sequences in the sample, most likely from *T. arcticus*, which shows > 90% similarity to both *S. alpinus* and *O. mykiss* for the 12S marker based on the unique sequences available in the reference database. The low number of unique sequences in GenBank for these species (three unique sequences for *O. mykiss*, two for *T. arcticus*, one for *S. alpinus*) suggests that both intra- and inter-species variation were not well captured, increasing the likelihood of misidentification, albeit in only one of three MiSeq runs.

Within the samples collected in the field, we observed a greater level of variation between independent MiSeq

runs. We suggest this is most likely due to variability in the amount of DNA obtained by the DNA extraction process, since multiple extractions were performed on each sample. Despite the differences in the number of reads found between Illumina MiSeq runs, however, the list of species identified based on the presence of sequence data matched almost exactly the list of species documented using traditional collection methods. Olds et al. (2016) compared depletion-based electrofishing and eDNA methods to determine species identities and richness. They found that eDNA-based estimates were on par with traditional depletion-based electrofishing methods with the added benefits of identification of additional species, reduced sampling effort, and reduced harassment of the fish. Our findings support this idea.

This study relied on single 2 L samples collected at different times and locations. The observed variability in the number of reads between samples and runs could be mitigated by increasing the sample volume, frequency of sample collection, and collection of replicate samples. Hydrological factors such as the residence time of water within the study system, spatial separation of species habitats, or species-specific differences in the rate of eDNA generation or decay, may also play a role in the detectability of certain species (Shogren et al. 2017; Tillotson et al. 2018; Sansom and Sasse 2017). As such, we suggest that multiple locations within the target water body be sampled, depending on the hydrology of the system, *i.e.* fewer sites for streams/rivers, more sites for larger and/or deeper lakes with no outlets.

The ability to design and multiplex a limited set of primers targeting multiple species in one Illumina MiSeq run exponentially increases the usefulness of an eDNA approach to environmental monitoring. Primer sets and reference databases can be customized to meet the needs of multispecies eDNA-based surveys at a variety of spatial scales and time-scales. The use of the Python/Biopython scripts to identify exact matches of a species in an eDNA sample to a custom reference database eliminates errors introduced by commercial bioinformatics software that looks for percent similarity and assigns costs for insertions/deletions. These scripts can also allow researchers to increase the lower threshold for acceptable identification from its current value of 0.01% and still avoid false positives. Combining locally-collected sequence data with data available from public repositories, by more broadly representing population-level sequence variation, should facilitate both improved specificity in primer design, as well as enhanced ability to uniquely identify related species in an eDNA-based survey.

Acknowledgements This research was funded by the National Park Service and the U. S. Geological Survey, Alaska Science Center. We thank Matthew Whitman (Bureau of Land Management) and Chris Arp (University of Alaska, Fairbanks) for the collection of environmental DNA samples from the Arctic Coastal Plain lakes, and for

providing fish survey data from those same lakes. We thank Andres Lopez (University of Alaska, Fairbanks) for providing access to the aquarium at the School of Fisheries and Ocean Sciences, and for input during the early stages of the project. We thank Randy Brown (U.S. Fish and Wildlife Service) for assistance in sorting out the genetics and taxonomy of Alaskan whitefish species. We thank Melanie Flamme (National Park Service) for her integral role in the development of the project and for providing critical assistance with obtaining funding. We thank Meg Gravley and Cody Dial (U. S. Geological Survey) for assistance with data collation during the early phases of the study, and Meg Gravley for assistance with graphics. Barbara Pierson (U. S. Geological Survey) assisted with data curation. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U. S. Government.

Data Accessibility Illumina MiSeq data can be found at NCBI BioProject PRJNA389325 <http://www.ncbi.nlm.nih.gov/bioproject/389325>.

References

- Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, Kjaer KH, ... Willersley E (2012) Metabarcoding of ‘dirt’ DNA from soil reflects vertebrate biodiversity. *Mol Ecol* 8:1966–1979. <https://doi.org/10.1111/j.1365-294X.2011.05261.x>
- Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* 10:57–59. <https://doi.org/10.1038/nmeth.2276>
- Brown RJ, Brown C, Braem NM, Carter WK III, Legere N, Slayton L (2012) Whitefish biology, distribution, and fisheries in the Yukon and Kuskokwim River drainages in Alaska: a synthesis of available information. US Fish and Wildlife Service, Alaska Fisheries Data Series Number 2012-4, Fairbanks, Alaska
- Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A et al (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 22(11):1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>
- Dejean T, Valentini A, Duparc A, Pellier-Cuit S, Pompanon F, Taberlet P, Miaud C (2011) Persistence of environmental DNA in freshwater ecosystems. *PLoS ONE* 6(8):e23398. <https://doi.org/10.1371/journal.pone.0023398>
- Eiler A, Lofgren A, Hjerne O, Norden S, Saetre P (2018) Environmental DNA (eDNA) detects the pool frog (*Pelophylax lessonae*) at times when traditional monitoring methods are insensitive. *Nat Sci Rep* 8:5452. <https://doi.org/10.1038/s41598-018-23740-5>
- Evans NT, Lamberti GA (2018) Freshwater fisheries assessment using environmental DNA: a primer on the method, its potential, and shortcoming as a conservation tool. *Fish Res* 197:60–66. <https://doi.org/10.1016/j.fishres.2017.09.013>
- Evans NT, Li Y, Renshaw MA, Olds BP, Deiner K, Turner CR, Jerde CL, Lodge DM, Lamberti GA, Pfrender ME (2017) Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Can J Fish Aquat Sci* 74(9):1–13
- Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biol Lett* 4(4):423–425. <https://doi.org/10.1098/rsbl.2008.0118>
- Foran DR (2006) Relative degradation of nuclear and mitochondrial DNA: an experimental approach. *J Forensic Sci* 51:766–770. <https://doi.org/10.1111/j.1556-4029.2006.00176.x>
- Goldberg CS, Pilliod DS, Arkle RS, Waits LP (2011) Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain Tailed Frogs and Idaho Giant Salamanders. *PLoS ONE* 6(7):e22746. <https://doi.org/10.1371/journal.pone.0022746>
- Greiman SE, Cook JA, Tkach VV, Hoberg EP, Menning SM, Hope AG, Sonsthagen SA, Talbot SL (2018) Museum metabarcoding: a novel method revealing gut helminth communities of small mammals across space and time. *Int J Parasitol* doi: <https://doi.org/10.1016/j.ijpara.2018.08.001>
- Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: a Next-Generation sequencing approach for biomonitoring application using river benthos. *PLoS ONE* 6(4):e17497. <https://doi.org/10.1371/journal.pone.0017497>
- Haynes TB, Rosenberger AE, Lindberg MS, Whitman M, Schmutz JA (2014) Patterns of lake occupancy by fish indicate different adaptations to life in a harsh Arctic environment. *Freshw Biol* 59(9):1884–1896. <https://doi.org/10.1111/fwb.12391>
- Haynes TB, Schmutz JA, Bromaghin JF, Iverson SJ, Padula VM, Rosenberger AE (2015) Diet of yellow-billed loons (*Gavia adamsii*) in Arctic lakes during the nesting season inferred from fatty acid analysis. *Polar Biol* 38(8):1239–1247. <https://doi.org/10.1007/s00300-015-1690-3>
- Hinlo R, Furlan E, Sutor L, Gleeson D (2017) Environmental DNA monitoring and management of invasive fish: comparison of eDNA and fyke netting. *Manag Biol Invasions* 8(1):89–100. <https://doi.org/10.3391/mbi.2017.8.1.09>
- Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, Burrige M et al (2008) Identifying Canadian freshwater fishes through DNA barcodes. *PLoS ONE* 3(6):e2490. <https://doi.org/10.1371/journal.pone.0002490>
- Jerde CL, Mahon AR, Chadderton WL, Lodge DM (2011) “Sight-unseen” detection of rare aquatic species using environmental DNA. *Conserv Lett* 4(2):150–157. <https://doi.org/10.1111/j.1755-263X.2010.00158.x>
- Koichiro T, Glen S, Daniel P, Alan F, Sudhir K (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Laramie MB, Pilliod DS, Goldberg CS (2015) Characterizing the distribution of an endangered salmonid using environmental DNA analysis. *Biol Conserv* 183:29–38. <https://doi.org/10.1016/j.biocn.2014.11.025>
- Laske SM, Haynes TB, Rosenberger AE, Koch JC, Wipfli MS, Whitman M, Zimmerman CE (2016) Surface water connectivity drives richness and composition of Arctic lake fish assemblages. *Freshw Biol*. <https://doi.org/10.1111/fwb.12769>
- Longmire JL, Maltbie M, Baker RJ (1997) Use of “lysis buffer” in DNA isolation and its implication for museum collections. *Occasional Papers Museum of Texas Tech University*, 163
- McDermid JL, Reist JD, Bodaly RA (2007) Phylogeography and post-glacial dispersal of whitefish (*Coregonus clupeaformis* complex) in northwestern North America. *Adv Limnol* 60:91–109
- McKelvey KS, Young MK, Knotek WL, Carim KJ, Wilcox TM, Padgett-Stewart TM, Schwartz MK (2016) Sampling large geographic areas for rare species using environmental DNA: a study of bull trout *Salvelinus confluentus* occupancy in western Montana. *J Fish Biol* 88:1215–1222. <https://doi.org/10.1111/jfb.12863>
- Mecklenburg CW, Mecklenburg TA, Thorsteinson LK (2002) Fishes of Alaska, 6th edn. American Fisheries Society, Bethesda
- Menning DM, Talbot SL (2018) Python scripts for bioinformatics, 2017. U.S. Geological Survey data release. <https://doi.org/10.5066/F74F1NZ4>
- Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson K-H, Koljalg U (2006) Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS ONE* 1:e59. <https://doi.org/10.1371/journal.pone.0000059>
- O’Rourke DH, Hayes MG, Carlyle SW (2000) Ancient DNA studies in physical anthropology. *Annu Rev Anthropol* 29:217–242. <https://doi.org/10.1146/annurev.anthro.29.1.217>

- Olds BP, Jerde CL, Renshaw MA, Li Y, Evans NT, Turner CR, ... Lamberti GA (2016) Estimating species richness using environmental DNA. *Ecol Evol* 6:4214–4226. <https://doi.org/10.1002/ece3.2186>
- Rodgers TW, Olson JR, Klobucar SL, Mock KE (2017) Quantitative PCR assays for detection of five arctic fish species: *Lota lota*, *Cottus cognatus*, *Salvelinus alpinus*, *Salvelinus malma*, and *Thymallus arcticus* from environmental DNA. *Conserv Genet Resour*. <https://doi.org/10.1007/s12686-017-0883-1>
- Sansom BJ, Sassoubre LM (2017) Environmental DNA (eDNA) shedding and decay rates to model freshwater mussel eDNA transport in a river. *Environ Sci Technol* 51:14244–14253. <https://doi.org/10.1021/acs.est.7b05199>
- Shen Y-Y, Chen X, Murphy RW (2013) Assessing DNA barcoding as a tool for species identification and data quality control. *PLoS ONE* 8:e57125. <https://doi.org/10.1371/journal.pone.0057125>
- Shogren AJ, Tank JL, Andruszkiewicz E, Olds B, Mahon AR, Jerde CL, Bolster D (2017) Controls on eDNA movement in streams: transport, retention, and resuspension. *Sci Rep* 7:5065. <https://doi.org/10.1038/s41598-017-05223-1>
- Spear SF, Groves JD, Williams LA, Waits LP (2015) Using environmental DNA methods to improve detectability in a hellbender (*Cryptobranchus alleganiensis*) monitoring program. *Biol Conserv* 183:38–45. <https://doi.org/10.1016/j.biocon.2014.11.016>
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol Ecol* 21:2045–2050. <https://doi.org/10.1111/j.1365-294X.05470>
- Thomsen PF, Willerslev E (2015) Environmental DNA—an emerging tool in conservation for monitoring past and present biodiversity. *Biol Conserv* 183:4–18. <https://doi.org/10.1016/j.biocon.2014.11.019>
- Tillotson MD, Kelly RP, Duda JJ, Hoy M, Kralj J, Quin TP (2018) Concentrations of environmental DNA (eDNA) reflect spawning salmon abundance at fine spatial and temporal scales. *Biol Conserv* 220:1–11. <https://doi.org/10.1016/j.biocon.2018.01.030>
- Valentini A, Pompanon F, Taberlet P (2009a) DNA barcoding for ecologists. *Trends Ecol Evol* 24(2):110–117. <https://doi.org/10.1016/j.tree.2008.09.011>
- Valentini A, Miquel C, Nawaz MA, Bellemain E, Coissac E, Pompanon F, ... Taberlet P (2009b) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Mol Ecol Resour* 9(1):51–60. <https://doi.org/10.1111/j.1755-0998.2008.02352.x>
- Van Rossum G (1995) Python Tutorial, Technical Report CS-R9526. Centrum voor Wiskunde en Informatica (CWI), Amsterdam
- Vences M, Kohler J (2008) Global diversity of amphibians (Amphibia) in freshwater. *Hydrobiologia* 595:569–580. <https://doi.org/10.1007/s10750-007-9032-2>
- Wegleitner B, Jerde C, Tucker A, Chadderton W, Mahon A (2015) Long duration, room temperature preservation of filtered eDNA samples. *Conserv Genet Resour* 7:789–791. <https://doi.org/10.1007/s12686-015-0483-x>
- Wittwer C, Stoll S, Strand D, Vralstad T, Nowak C, Thines M (2018) eDNA-based crayfish plague monitoring is superior to conventional trap-based assessments in year-round detection probability. *Hydrobiologia* 807(1):87–97. <https://doi.org/10.1007/s10750-017-3408-8>
- Yamamoto S, Masuda R, Sato Y, Sado T, Araki H, Kondoh M, ... Miya M (2017) Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Sci Rep* 7:40368. <https://doi.org/10.1038/srep40368>