Monitoring Harbour Fish Communities in Halifax, Nova Scotia using Environmental DNA

Prepared for:

Nick Jeffery
Aquatic Science Biologist
Fisheries and Oceans Canada
Nick.Jeffery@dfo-mpo.gc.ca

RPT037-OP01

July 15, 2022

Centre for Environmental Genomics Applications

Centre for Environmental Genomics Applications

eDNAtec Inc.

103-14 International Pl.
St. John's NL A1A 0R6
CANADA

T: +1 (709) 576-2550

services@ednatec.com
www.cegacanada.com

This page is intentionally left blank

The correct citation for this report is:

Center for Environmental Genomics Applications, Monitoring Harbour Fish Communities in Halifax, Nova Scotia using Environmental DNA, CEGA RPT037-OP01, 2022-07-15.

Project Team:

Mehrdad Hajibabaei, MH (Director)

Nicole Fahner, NF (Associate Director)

Eman Elbakry, EE (Research Technologist)

Muneesh Kaushal, MK (Bioinformatician)

Jasmine Yeung, JY (Junior Ecologist)

REVISION HISTORY

VERSION	**NAME**	**COMPANY**	**DATE OF ISSUE**	**COMMENTS**
R1	Lesley Berghuis	CEGA	2022-07-15	First release

DISTRIBUTION LIST

NAME	**COMPANY**	**NUMBER OF COPIES / FORMAT**
Mehrdad Hajibabaei	CEGA	1 / electronic [pdf]
Nicole Fahner	CEGA	1 / electronic [pdf]
Nick Jeffery	DFO	1/electronic [pdf]

TABLE OF CONTENTS

LIST OF TABLES

1 Executive Summary

Fisheries and Oceans Canada (DFO) is using environmental genomics for harbour monitoring in Halifax, NS. In 2021, DFO collected environmental water samples from harbours in Halifax and submitted them to CEGA for laboratory analysis. 100 Sterivex filters (96 environmental samples and 4 field blanks) and six DNA extracts were submitted and received for analysis. DNA was extracted from all filters and two DNA markers (12S MiFishU and 16S from McInnes et al. 2017) were amplified and then sequenced on an Illumina NovaSeq 6000 instrument. DNA extraction and sequencing was successful for all samples. Raw sequencing results were demultiplexed and trimmed and provided to DFO for further bioinformatics analysis.

2 Background

Environmental genomics is a novel approach to biodiversity characterization that does not require collection of whole biological specimens but instead relies on recovery and analysis of DNA from the physical environment in which they live (e.g., water, soil, sediment, etc.). This environmental DNA (eDNA) is released from organisms through various mechanisms including cell shedding and the excretion of various bodily fluids and feces [1–3]. DNA is isolated from samples, biodiversity informative regions of DNA are amplified through polymerase chain reaction (PCR), and then these regions are sequenced using high-throughput genomic sequencing platforms.

There are several advantages to environmental genomics compared to the conventional approach to biodiversity characterization. It is non-invasive and does not require the capture of a whole biological specimen. Very small samples of water (~250 mL) or sediment (~5 g) are required for a biodiversity assessment because, unlike conventional surveys, organisms at all trophic levels from algae to large mammals can be detected from the same environmental sample. The technique is highly sensitive, which means rare or elusive species, including endangered and invasive species, can be detected through their DNA in addition to the more common species. It is cost effective with time savings during field sampling and sample analysis. Furthermore, high-throughput DNA sequencing technology allows for simultaneous analysis of a large set of samples.

Because environmental genomics is a novel approach to biodiversity characterization, it is not widely applied yet. However, in recent years, there has been a steady increase in its application to various environmental characterization efforts [4–6]. While the sensitivity of environmental genomics is one of the many appeals of this technology, it is also one of its potential limitations as contaminants of biological nature can also be detected within the environmental sample. To manage this, many precautions are taken during the acquisition, preservation, and processing of samples to limit the introduction of contaminants (e.g. collection of field blanks as a negative control).

For this project, Fisheries and Oceans Canada (DFO) is using environmental genomics for harbour monitoring, particularly of fish biodiversity, in Halifax, Canada.

3 Methods

3.1 Sample Submission

1L environmental water samples were collected and filtered by DFO from harbours in Halifax in August 2021 and submitted to CEGA for lab analysis. CEGA received 100 Sterivex filters (96 environmental samples and 4 field blanks) preserved in Longmire's buffer and six DNA extracts. All samples were received frozen and were stored at -80°C until processing.

3.2 Sample Processing

DNA was extracted from filters for each sample and field blank and quantified for all samples, including those received as DNA extracts (Tables 1-3). Two DNA markers for fish biodiversity, 12S MiFishU and the 16S marker from McInnes et al. (2017)[7], were amplified in all samples and blanks. Additional negative controls were added during extraction and PCR stages, which were carried through subsequent stages. All samples, including field and lab negative controls, were sequenced on an Illumina NovaSeq 6000 instrument using a 2 x 150 cycle SP kit and a minimum target depth of 1,000,000 reads per sample per marker.

3.3 Data Analysis

Minimal processing was done to produce the fastq files from the raw sequencer data. Demultiplexing was performed using Illumina's BCL Convert software (version 3.10.5). Primers were then trimmed from the forward and reverse reads using cutadapt v4.1. The maximum error rate was set to 0.07 with a minimum overlap equal to 13 bases. Read pairs in which the primer was missing from either the forward or reverse read were discarded.

Table 1. List of water eDNA filters received for analysis and the DNA extract concentration.

Sample Name	[DNA] (pg/µL)
485525	621
485526	1650
485532	383
485533	1995
485539	741
485540	1213
485543	707
485544	1049
485545	807
485546	1893
485547	523
485548	2082
485563	1792
485564	1882
485568	334
485569	2161
485573	822
485574	1821
485589	1135
485590	1644
485594	1928
485595	1063
485599	3961
485600	1714
485603	500
485604	317
485605	572
485606	1722
485607	325
485608	1922
485610	454
485611	1456
485612	902
485613	1716
485614	781
485615	1705
485617	1867

Sample Name	[DNA] (pg/µL)
485618	2984
485619	1277
485620	1609
485621	1251
485622	2455
485624	721
485625	793
485626	1254
485627	1745
485628	454
485629	1375
485631	781
485632	3257
485633	905
485634	3719
485635	845
485636	4361
485638	758
485639	5165
485640	741
485641	4376
485642	509
485643	3666
485645	1098
485646	2665
485647	1133
485648	2817
485649	1375
485650	3687
485652	1176
485653	3852
485654	1358
485655	4887
485656	1699
485657	4691
485659	615
485660	911

Sample Name	[DNA] (pg/µL)
485661	454
485662	1517
485663	598
485664	756
485666	388
485667	1896
485668	1006
485669	1801
485670	331
485671	2353
485673	919
485674	2762
485675	948
485676	2548
485677	603
485678	4198
485680	1285
485681	2504
485682	635
485683	2820
485684	1511
485685	2750

Table 2.List of water filter field blanks received for analysis and the DNA extract concentrations. BQL indicates that the DNA in that sample was below detection limit.

Sample Name	[DNA] (pg/µL)
485609	22
485658	BQL
485686	BQL
485549	BQL

Table 3. List of DNA extracts received for analysis and their concentrations. BQL indicates that the DNA in that sample was below detection limit.

Sample Name	[DNA] (pg/µL)
DFOGUL21_41102	51
DFOGUL21_41103	91
DFOGUL21_41104	265
DFOGUL21_41105	BQL
DFOGUL21_41106	1138
DFOGUL21_41107	279

4 Results

The mean DNA concentration recovered from the submitted filters was 1664 pg/µL (range: 317 – 5165 pg/µL, Table *1*) and all submitted field blanks had DNA concentrations below quantifiable levels, except for one which had a DNA concentration of 22 pg/µL (Table *2*). The mean DNA concentration of the submitted DNA extracts was 365 pg/µL (range: 51 – 1138 pg/µL with one sample below the quantification limit (Table 3)). There was an average of 1,321,008 sequencing reads (range: 39,969 – 5,860,704 reads) per sample per marker, excluding lab and field negative controls.

In summary, DNA was extracted and sequenced successfully for all samples. Files with raw sequence data for further bioinformatics analysis will be provided to DFO.

5 References

1. Baird, D. J. & Hajibabaei, M. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Mol. Ecol.* **21**, 2039–2044 (2012).
2. Taberlet, P., Coissac, E., Hajibabaei, M. & Rieseberg, L. H. Environmental DNA. *Mol. Ecol.* **21**, 1789–1793 (2012).
3. Thomsen, P. F. *et al.* Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE* **7**, e41732 (2012).
4. Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C. & Baird, D. J. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE* **6**, e17497 (2011).
5. Leray, M. & Knowlton, N. Censusing marine eukaryotic diversity in the twenty-first century. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150331 (2016).
6. Gibson, J. F. *et al.* Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. *PLOS ONE* **10**, e0138432 (2015).
7. McInnes, J. C. *et al.* DNA Metabarcoding as a Marine Conservation and Management Tool: A Circumpolar Examination of Fishery Discards in the Diet of Threatened Albatrosses. *Front. Mar. Sci.* **4**, 277 (2017).