Environmental DNA Analysis of Water Samples from Musquash Estuary Marine Protected Area

Prepared for:

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

RPT063-HE03

June 13, 2023

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

REVISION HISTORY

VERSION	**NAME**	**COMPANY**	**DATE OF ISSUE**	**COMMENTS**
R1	Jasmine Yeung and Ariel Lisogorsky	CEGA	June 13, 2023	First release

DISTRIBUTION LIST

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

The correct citation for this report is:

Center for Environmental Genomics Applications, Environmental DNA Analysis of Water Samples from Musquash Estuary Marine Protected Area, CEGA Report RPT063-HE03, 2023-06-13.

TABLE OF CONTENTS

1 Executive Summary 1
2 Background 2
3 Methods 3
3.1 Sample Submission 3
3.2 Sample Processing 3
3.3 Data Analysis 3
4 Results 7
5 References 8
6 Appendix A 9

LIST OF TABLES

Table 1. List of water eDNA samples collected from Musquash Estuary Marine Protected Area in September 2022 that were submitted to CEGA for analysis with their DNA concentrations after extraction. 4
Table 2. List of lab and field blanks samples collected as part of the Musquash Estuary Marine Protected Area sampling in August 2022 that were submitted to CEGA for analysis with their DNA concentrations after extraction. BQL indicates the DNA concentration was below the quantifiable limit (<0.1 ng/µL)..... 6

Table A1 List of all supplementary files to be shared with the report. 9

1 Executive Summary

Fisheries and Oceans Canada (DFO) is using environmental genomics for fish and general marine biodiversity of the Musquash Estuary Marine Protected Area (MPA). In September 2022, DFO collected water samples from the MPA and filtered these samples in December 2022. Filters were then submitted to the Centre for Environmental Genomics Applications (CEGA) for laboratory analysis and DNA was successfully extracted from the filters. Following a DNA metabarcoding approach, three DNA markers for fish and general marine biodiversity were amplified from all samples and then sequenced on an Illumina NovaSeq 6000 instrument. Sequencing was successful for all samples. Raw sequencing results were demultiplexed 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. The resulting DNA sequences are filtered through a variety of quality control and assurance steps and then compared to publicly available databases (e.g., GenBank) where the genomic information for known specimens has been deposited. When a match is found, a taxonomic identification can be assigned to a DNA sequence from an environmental sample[1–3].

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). In addition, stringent quality control measures during data analysis can limit false positives.

For this project, Fisheries and Oceans Canada (DFO) is using environmental genomics to characterize fish and general marine biodiversity in the Musquash Estuary Marine Protected Area (MPA). Water samples were collected from the MPA and filtered by DFO. Samples were then sent to the Centre for Environmental Genomics Applications (CEGA) for metabarcoding library preparation and sequencing.

3 Methods

3.1 Sample Submission

Surface and bottom triplicate water eDNA samples were collected from the Musquash Estuary MPA on September 12 and 13, 2022 and a field negative control was collected each sampling day. Water samples were frozen and then filtered by DFO on December 12, 2022, and two filtration negative controls were generated. Filters were preserved in Longmire's buffer and frozen at -20°C but thawed during transport to CEGA. To avoid further freeze-thaw cycles, the filters were stored at 4°C until processing. In total, CEGA received 81 environmental samples and 4 negative controls as filters.

3.2 Sample Processing

DNA was extracted from all filters and the DNA in each sample was quantified for sample and field blanks. Three DNA markers from three gene regions (12S for fish diversity, CO1 and 18S for general biodiversity and invertebrates) were amplified in all samples and blanks. Additional negative controls were added during extraction and PCR stages and all blanks were carried through to sequencing. All samples, including field and lab negative controls, were sequenced on an Illumina NovaSeq 6000 instrument using a SP500 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 by sample was performed using Illumina's BCL Convert software (version 4.0.3) and primer sequences were trimmed from the sequencing files using cutadapt (version 4.3). FastQC (version 0.12.1) was used to assess the quality of sequences.

Table 1. List of water eDNA samples collected from Musquash Estuary Marine Protected Area in September 2022 that were submitted to CEGA for analysis with their DNA concentrations after extraction.

Sample Name	DNA Concentration (ng/µL)
M06S2	2.54
M04S3	2.33
M04S2	3.26
M04S1	2.20
M02S1	5.64
M01S3	3.59
M01S2	3.44
M01S1	3.73
M11S2	3.84
M11S1	2.89
M08B1	1.18
M08S3	2.36
M08S2	4.06
M07S2	3.06
M07S1	3.20
M06S3	3.02
M19S2	2.94
M19S1	2.45
M16S3	1.51
M16S2	2.63
M15S3	2.72
M15S2	3.54
M15S1	2.04
M11S3	3.90
M24S1	1.33
M22S3	2.38
M22S2	3.84
M22S1	3.56
M20B3	2.00
M20B2	2.53
M20B1	0.47
M19S3	2.46
M30S1	1.59

Sample Name	DNA Concentration (ng/µL)
M28S3	0.42
M28S2	2.39
M28S1	1.42
M26S2	2.34
M26S1	2.54
M24S3	1.59
M24S2	1.37
M32B2	2.10
M32B1	1.53
M32S3	2.04
M32S2	1.64
M30B3	1.88
M30B2	2.84
M30B1	1.94
M30S2	0.87
2B03S3	1.60
2B03S2	1.32
2B03S1	1.19
2B02S2	1.85
2B02S1	2.16
2B01S2	1.30
2B01S1	2.61
M32B3	1.85
301B2	0.90
301B1	1.23
301S3	2.16
301S2	1.93
301S1	1.91
2B03B3	1.25
2B03B2	1.44
2B03B1	1.03
2A01S1	0.86
303S3	0.87

Sample Name	DNA Concentration (ng/µL)
303S2	0.52
303S1	1.12
302S3	1.73
302S2	0.37
302S1	1.76
301B3	2.65
2A12B2	1.71
2A12B1	2.17
2A12S3	1.22
2A12S2	1.63
2A12S1	1.62
2A10S3	0.59
2A01S3	1.21
2A01S2	1.61
2A12B3	1.43

Table 2. List of lab and field blanks samples collected as part of the Musquash Estuary Marine Protected Area sampling in August 2022 that were submitted to CEGA for analysis with their DNA concentrations after extraction. BQL indicates the DNA concentration was below the quantifiable limit (<0.1 ng/µL).

Sample Name	DNA Concentration (ng/µL)
LB02	BQL
LB01	BQL
B0913	BQL
B0912	BQL

4 Results

DNA was successfully extracted from 81 water samples and four blanks collected from Musquash Estuary Marine Protected Area in 2022. The mean DNA concentration recovered from the submitted filters was 2.07 ng/µL (range: 0.37 – 5.64 ng/µL, **Table 1**) and all submitted field blanks had DNA concentrations below quantifiable levels (<0.1 ng/µL, **Table 2**).

A total of 532,525,166 paired-end reads were successfully sequenced markers with an average 89% bases with quality scores ≥Q30. After demultiplexing by sample, samples yielded an average of 3,257,377 paired end reads across three markers. In summary, DNA was amplified and sequenced successfully for three DNA markers from all samples. Raw sequence data are provided via FTP site alongside this report for further bioinformatic analysis.

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. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).

8. Freeland, J. The importance of molecular markers and primer design when characterizing biodiversity from environmental DNA (eDNA). *Genome* **60**, 52 (2016).

6 Appendix A

Table A1 List of all supplementary files to be shared with the report.

Name	Filename	Description
Sequence Data	Multiple files	Compressed FASTQ files (one file per sample per marker per forward and reverse read) where Illumina adapters, indexes, and primer sequences have been trimmed. No other manipulation has been performed.