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

Prepared for:

Andrew Cooper, PhD.
Research Scientist
Fisheries and Oceans Canada
Andrew.Cooper@dfo-mpo.gc.ca

RPT089-HE04

March 28, 2024

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:

Centre for Environmental Genomics Applications, Environmental DNA Analysis of Water Samples from Musquash Estuary Marine Protected Area, RPT089-HE04, 2024-03-28.

REVISION HISTORY

VERSION	NAME	COMPANY	DATE OF ISSUE	COMMENTS
R1	Beverly McClenaghan	CEGA	March 28, 2024	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]
Andrew Cooper	Fisheries and Oceans Canada	1 / electronic [pdf]

TABLE OF CONTENTS

LIST OF TABLES

1 Executive Summary

Fisheries and Oceans Canada (DFO) is using environmental DNA to assess fish and general marine biodiversity of the Musquash Estuary Marine Protected Area (MPA). In June and September 2023, DFO collected water samples for environmental genomics analysis from the MPA and submitted filters to the Centre for Environmental Genomics Applications for laboratory analysis. Following a DNA metabarcoding approach, two DNA markers for fish and general marine biodiversity were amplified from all samples and then sequenced on an Illumina NovaSeq 6000 instrument. DNA extraction and sequencing was successful for all samples. Sequencing results were demultiplexed and primers were trimmed then data were 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

Water eDNA samples were collected and filtered from the Musquash Estuary MPA in June and September 2023 by DFO. Two field negative controls and two lab filtration negative controls were also collected. Samples were filtered on self-preserving Smith-Root filters, which were stored at room temperature for transport to CEGA. At CEGA, filters were stored at -80°C until processing. In total, CEGA received 82 environmental samples and four negative controls.

3.2 Sample Processing

DNA was extracted from all filters and the DNA was quantified for each sample (**Table *1***) and negative control (**Table *2***). Two DNA markers from two gene regions (COI and 12S) 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 were sequenced on an Illumina NovaSeq 6000 instrument with 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 4.2.7) and primer sequences were trimmed from the sequencing files using cutadapt (version 4.6). FastQC (version 0.12.1) was used to assess the quality of sequences.

Table 1 List of water eDNA samples collected from Musquash Estuary MPA and 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] ng/µL
424082	2.87
424081	3.48
424080	4.34
424079	*BQL*
424078	8.21
424077	4.23
424076	6.83
424075	6.48
424074	7.63
424073	4.85
424072	8.15
424071	11.37
424070	13.15
424069	5.81
424068	5.36
424067	7.53
424066	13.71
424065	6.57
424064	7.42
424063	5.84
424062	9.19
424061	11.87
424060	6.44
424059	3.65
424058	5.86
424057	2.89
424056	3.03
424055	2.32
424054	3.48
424053	3.29
424052	2.95
424051	3.46
424050	2.26
424049	2.95

Sample Name	[DNA] ng/µL
424048	2.17
424047	2.83
424046	2.78
424045	8.85
424044	7.11
424043	10.30
424042	2.34
424041	3.82
424040	3.52
424039	3.69
424038	2.79
424037	2.38
424036	3.39
424035	3.93
424034	2.41
424033	3.93
424032	5.98
424031	4.96
424030	3.34
424029	3.04
424028	2.58
424027	2.54
424026	3.34
424025	2.32
424024	3.24
424023	2.12
424022	1.98
424021	2.56
424020	2.97
424019	3.29
424018	2.07
424017	3.15
424016	3.39
424015	3.11

Sample Name	[DNA] ng/µL
424014	2.09
424013	3.27
424012	4.39
424011	3.89
424010	4.00
424009	2.86
424008	2.22

Sample Name	[DNA] ng/µL
424007	3.32
424006	3.12
424005	4.12
424004	2.75
424003	2.22
424002	2.91
424001	3.37

Table 2 List of negative controls submitted to CEGA for analysis, with their DNA concentrations after extraction.

Sample Name	Negative Control Step	[DNA] ng/µL
424083	Field	3.62
424084	Field	4.21
424085	Filtration	5.48
424086	Filtration	6.17

4 Results

DNA was extracted from all 82 seawater samples and four field blanks collected from the Musquash Estuary MPA in 2023. DNA was quantified from all samples after extraction. One sample did not have quantifiable DNA (sample name: 424079). The median quantifiable DNA concentration recovered from the submitted filters was 3.39 ng/µL (range: 1.98 – 13.71 ng/µL, **Table 1**). All submitted field and filtration blanks had quantifiable DNA concentrations (3.63 – 6.17 ng/µL, median: 4.85 ng/µL; **Table 2**). All CEGA internal lab blanks were below the quantifiable limit for DNA.

A total of 304 million paired-end reads were successfully sequenced across two markers with an average of 88% of bases with quality scores ≥Q30 across all samples and markers. After demultiplexing by sample, samples yielded an average of 1.5 million paired end reads per sample across the two markers. In summary, DNA was amplified and sequenced successfully for two DNA markers from all samples. Sequence data are provided via FTP site alongside this report for further bioinformatic analysis.

Sequences were recovered from field and filtration blanks. A preliminary analysis shows sequences largely attributed to *Clupea harengus*, *Microgadus tomcod*, *Anguilla rostrata*, *Scomber scombrus*, and *Merluccius bilinearis*. No bioinformatics analysis was requested, therefore CEGA did not perform any analytical steps to account for external contamination, cross-contamination, or index hopping. Appropriate contamination mitigation steps should be taken with the data prior to ecological analysis and interpretation.

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).

6 Appendix A

Table A1 List 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.