Dataset Description: Environmental DNA Sequences and Extractions from Wells, Great Bay, Hudson, Apalachicola, South Slough, and He'eia National Estuarine Research Reserves

This document provides detailed information about three datasets that were generated through the 2017-2020 collaborative research project *New Technology for Old Problems: Developing DNA Methods to Monitor Invasive Species and Biodiversity in Estuaries*. The project was supported by the National Estuarine Research Reserve System (NERRS) Science Collaborative, which is funded by the National Oceanic and Atmospheric Administration. All Science Collaborative supported projects that collect new data adhere to federal data sharing and archiving requirements.

Data access and archival: All genetic datasets from this project have been archived and made publicly available at the National Center for Biotechnology Information, BioProject ID: PRJNA667067, at: https://www.ncbi.nlm.nih.gov/search/all/?term=PRJNA667067 Summary tables of processed eDNA data and associated fish species identifications are available at the project webpage: estuarydna.org.

List of Project Datasets:
Three datasets related to eDNA sequences and extractions from six National Estuarine Research Reserves are described in this document.

1. 12S sequence data and summary of fish species
2. 18S sequence data
3. Archived DNA extractions from water samples from 6 Reserves

About the Associated Project

Project title: Developing DNA Methods to Monitor Invasive Species and Biodiversity in Estuaries
Name of reserves involved in the project: Apalachicola (FL), Great Bay (NH), He'eia (HI), Hudson River (NY), South Slough (OR), and Wells (ME) National Estuarine Research Reserves
Project period: 2017 to 2020
Science Collaborative project page: www.nerrssciencecollaborative.org/project/Watts17
Project lead and contact information: Alison Watts, University of New Hampshire, Email: alison.watts@unh.edu

Purpose: Biological monitoring programs are essential foundations for effective management of estuaries and coasts, but they can be expensive to conduct and methods such as the capture of living organisms may be traumatic for some target species. Advancements in DNA methods now make it possible to identify the organisms in an area by the DNA they leave behind, but there are relatively few proven sampling protocols and example datasets from estuaries. This project established a Collaborative Learning Community of Practice around Environmental DNA (eDNA) sampling within the National Estuarine Research Reserve system and explored the potential for eDNA to support estuarine management. Scientists and staff from six reserves worked

collaboratively with researchers, resource agencies, and a technical advisory team to design and implement eDNA-based monitoring protocols for specific issues and species of interest in estuarine ecosystems.

Abstract: This project tested the use of eDNA methods to identify fish communities, assess biodiversity, and detect invasive crabs in different types of estuaries and connected streams. Samples were collected at Apalachicola, Great Bay, He'eai, Hudson, South Slough and Wells Reserves in 2018 and 2019. The project deployed two methods to analyze DNA in water samples: metabarcoding that identifies a wide range of species, and digital droplet Polymerase Chain Reaction (ddPCR) that identifies a single target organism. The team conducted baseline community assessments for fish at five reserve sites and conducted a survey of anadromous fish in a tributary stream at the sixth reserve. The project found that fish community and biodiversity assessments are well suited to eDNA applications, while invasive crabs are much harder to detect because they do not shed much DNA. The project developed protocols and recommendations for the collection, filtering, and extraction of eDNA samples at estuarine sites, and provided information which will support the design of sampling programs for fish communities.

About the Project Datasets

1. 12S Sequence Data and Summary of Fish Species

General description of data: This data set includes amplicon sequences generated from water samples collected at six estuary reserves. The samples were filtered at each site, then extracted at the University of New Hampshire, and sequenced using a 12S primer designed for fish detection (Miya et al 2015). Data files contain the sequences identified, and curated list of fish species associated with each sequence.

More about the data: Associated metadata includes site location (lat/long), sample volume, and other relevant environmental information (for instance tide level at some sites).

Data collection period: June 2018 to October 2019

Geographic extent: Samples were collected at Wells, Great Bay, Hudson, Apalachicola, South Slough and He'eia National Estuarine Research Reserves (see Figure 1). Between 4 and 17 sites were sampled at each reserve, including tributary streams at some locations. Latitude and longitude for each site are provided with the metadata in summary tables.

File format: Excel

Summary data tables: Summary data tables containing MiFish sequences and fish identification for five reserves are available on the project website: estuarydna.org. These data tables summarize the results of baseline fish community assessments at each reserve and address questions specific to that location.

Data Access and Archival: Genetic sequence data have been archived and made publicly available at the National Center for Biotechnology Information, BioProject ID: PRJNA667067.

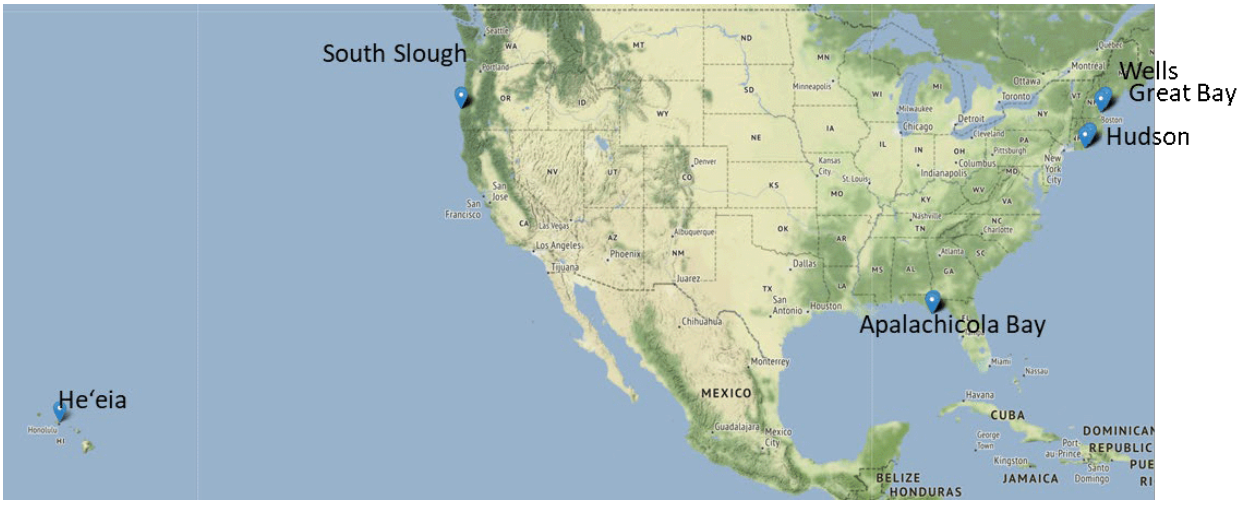

Figure 1: Map of the six National Estuarine Research Reserves where eDNA were collected

2. 18S Sequence Data

General description of data: This data set includes amplicon sequences generated from water samples collected at four estuary reserves. The samples were filtered at each site, then extracted at the University of New Hampshire, and sequenced using the Earth Microbiome 18S primer for eukaryote species.

More about the data: Associated metadata includes site location (lat/long), sample volume.

Data collection period: June 2019 to October 2019

Geographic extent: Samples were collected at Great Bay, Apalachicola, South Slough and He'eia National Estuarine Research Reserves. Between 4 and 12 sites were sampled at each reserve.

File format: Excel

Data Access and Archival: Genetic sequence data have been archived and made publicly available at the National Center for Biotechnology Information, BioProject ID: PRJNA667067.

3. Archived DNA Extractions from Water Samples from 6 Reserves

General description of data: This data set includes physical DNA samples extracted from water samples collected at six estuary reserves. The samples were filtered at each site, then extracted at the University of New Hampshire.

More about the data: Samples were extracted from 1.5um glass fiber filters using a QIAcube Connect system (QIAGEN®, Hilden, Germany) and following the QIAamp DNA Mini Kit protocol. All samples have been run for at least one set of species (fish or eukoryotes). Aliquots of the remaining sample extraction can be requested for analysis for other parameters, or for duplicate analyses. Metadata associated with each sample is included in the summary tables in Dataset 1: 12S Sequence Data and Summary of Fish Species.

Data collection period: June 2018 to October 2019

Geographic extent: Samples were collected at Wells, Great Bay, Hudson, Apalachicola, South Slough and He'eia National Estuarine Research Reserves (Figure 1) Between 4 and 17 sites were sampled at each location, including tributary streams at some sites. Latitude and longitude for each site are provided with the metadata.

Format: Samples are stored at -80C at the University of New Hampshire. Up to 50ul of sample are stored in each well of 12x8 PCR plates. Aliquots of 10ul can be obtained by request.

Data Access: Requests to uses these extractions should be made to the project lead:
Alison Watts
University of New Hampshire
Email: alison.watts@unh.edu