

# Amplicon sequence variants (ASVs) recovered from samples and their related identification as Pseudo-nitzschia taxa and the methods used

**Website:** <https://www.bco-dmo.org/dataset/847469>

**Data Type:** Cruise Results

**Version:** 1

**Version Date:** 2021-04-05

## Project

» [RII Track-1: Rhode Island Consortium for Coastal Ecology Assessment, Innovation, and Modeling](#) (C-AIM)

| Contributors                        | Affiliation   | Role                      |
|-------------------------------------|---|---------------------------|
| <a href="#">Jenkins, Bethany D.</a> | University of Rhode Island (URI)                    | Principal Investigator    |
| <a href="#">Bertin, Matthew</a>     | University of Rhode Island (URI)                    | Co-Principal Investigator |
| <a href="#">Sterling, Alexa</a>     | University of Rhode Island (URI)                    | Contact                   |
| <a href="#">Copley, Nancy</a>       | Woods Hole Oceanographic Institution (WHOI BCO-DMO) | BCO-DMO Data Manager      |

## Abstract

This dataset is related to approximately weekly sampling of Narragansett Bay, RI in tandem with the University of Rhode Island (URI) Graduate School of Oceanography (GSO) Long-Term Plankton Time Series (LTPTS) and Fish Trawl Survey to examine species assemblages and toxicity of the diatom genus Pseudo-nitzschia spp. This dataset includes the amplicon sequence variants (ASVs) recovered from samples and their related identification as Pseudo-nitzschia taxa and the methods used related to the Sterling et al manuscript. These data are connected to NCBI Bioproject PRJNA690940 & GenBank Accession Numbers MW447658 – MW447770, which will be released January 2025 or when the associated manuscript is published, whichever occurs first.

## Table of Contents

- [Coverage](#)
- [Dataset Description](#)
  - [Acquisition Description](#)
  - [Processing Description](#)
- [Related Publications](#)
- [Parameters](#)
- [Instruments](#)
- [Deployments](#)
- [Project Information](#)
- [Funding](#)

## Coverage

**Spatial Extent:** N:41.6716 E:-70.8626 S:40.206 W:-71.42

**Temporal Extent:** 2016-09-26 - 2019-11-25

## Acquisition Description

For most samples, plankton biomass for Pseudo-nitzschia DNA identification was collected by passing an

average of 270 mL of surface seawater with a peristaltic pump across a 25 mm 5.0 mm polyester membrane filter (Sterlitech, Kent, WA, USA). Widths of some *Pseudo-nitzschia* spp. are < 5.0 mm (Lelong et al. 2012), but this size pore likely captured horizontally orientated cells and chains of cells, and was consistent with pore size used to examine toxicity. Filters were flash frozen in liquid nitrogen and stored at -80 °C until extraction. DNA was extracted using a modified version of the DNeasy Plant DNA extraction kit (Qiagen, Germantown, MD, USA) with an added bead beating step for 1 minute and QIA-Shredder column (Qiagen, Germantown, MD, USA) as reported in Chappell et al. 2019. Additionally, DNA was eluted in 30 µL with a second elution step of either 30 or 15 µL to maximize DNA yield. DNA was assessed for quality with a Nanodrop spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) and quantified using a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) with the Broad Range dsDNA and High Sensitivity dsDNA kits (Thermo Fisher Scientific Inc., Waltham, MA, USA). DNA yields reported by the Qubit ranged from below the limit of detection to 26.5, with an average of 2.0 ng DNA / mL eluent. Long-Term Plankton Time Series (LTPTS) samples from October 2016 and March 2017 had an average of 300 mL surface seawater passed over a 25 mm 0.2 mm filter, were extracted following existing LTPTS methods of DNA extraction using the DNeasy Blood and Tissue Kit (Qiagen, Germantown, MD, USA) with an added bead beating step (Canesi and Rynearson 2016), and yielded average 0.9 ng DNA / mL eluent as measured by the Qubit. Net tow samples had 50 mL of concentrate was passed across a 0.22 µm pore size Sterivex filter unit (MilliporeSigma, Burlington, MA, USA), and were extracted with the same modified DNeasy Plant DNA extraction protocol as above, with 4x volumes of AP1 buffer and RNase A and beads added to the unit to account for the larger sample surface area, extraction occurring within the capped unit itself to maximize yield, and then the lysate removed with a sterile syringe and subsequent steps with adjusted volumes as appropriate. As expected, DNA yields were higher from the Sterivex units ranging from 2.4 – 54.0 ng DNA / mL eluent with an average of 13.7 ng DNA/ mL elution as measured by the Qubit. For the March 13, 2017 NBay samples, 125 mL of surface seawater was passed across a HV filter and extracted with the DNeasy Plant DNA extraction kit with scissors and no beads. As measured by the Qubit, the average DNA yield was 3.7 ng DNA / mL eluent. A negative control sample was prepared of a blank 25 mm 5.0 mm polyester membrane filter using extraction reagents which had no detectable DNA using the Qubit. There were two positive controls of mock communities comprised of two known *Pseudo-nitzschia* species from monocultures. The two *Pseudo-nitzschia* cultures were *P. subcurvata* collected from the Southern Ocean and *P. pungens* isolated from NBay (provided by J. Rines). One positive control was made by combining equal concentrations of extracted DNA with 1.0 ng DNA of each culture. The second positive control was created of equal cell abundance estimated to be captured onto the filters of the cultures prior to extraction. These negative and positive controls were prepared for sequencing and sequenced on the same plate as the other environmental samples.

The ITS1 has been targeted for amplification and analysis by ARISA previously for *Pseudo-nitzschia* identification in environmental samples (Hubbard, Rocap, and Armbrust 2008). A comparison of ITS1 appears to be much less conserved and is divergent enough across *Pseudo-nitzschia* that 41 different species can be identified using existing public sequencing data. The primers to target the ITS1 region of *Pseudo-nitzschia* used this existing forward primer sequence of the ITS1 region for eukaryotes: TCCGTAGGTGAACCTGCGG (White et al. 1990) and a custom reverse primer designed using 132 *Pseudo-nitzschia* ITS1 sequences from the NCBI nucleotide database (downloaded on 4/3/2019) from this nucleotide search: ((*Pseudo-nitzschia*[Organism]) AND internal transcribed spacer[Title]) NOT uncultured): CATCCACCGCTGAAAGTTGTAA. This reverse primer targets a conserved region in the 5.8S. All primer sequences are reported from 5' – 3'. MiSeq adapter sequences were added to the beginning of the primer sequences for these full sequences used in this study: forward primer TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCCGTAGGTGAACCTGCGG and reverse primer GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATCCACCGCTGAAAGTTGTAA. When checking the specificity of these primers using the NCBI nt database, it became known that sequences beyond *Pseudo-nitzschia* would also be amplified in this study including other diatoms and dinoflagellates; however, the large number of sequencing reads recovered on the MiSeq platform would circumvent this non-specific characteristic of the primers.

The accession numbers of the sequences used in this primer design are reported in Table S2 of Sterling et al. (in prep), along with a summary of *Pseudo-nitzschia* species expected to amplify with these based on the in silico design. The expected ranges for PCR products were from 235 – 370 bp as the size of the ITS1 region differs for some *Pseudo-nitzschia* taxa. Primers (Integrated DNA Technologies, Coralville, IA, USA)

were HPLC purified, resuspended in 1x Tris-Acetate-EDTA (TAE) buffer, and then working stocks created in diethylpyrocarbonate (DEPC)-treated H<sub>2</sub>O. About 4 ng of extracted DNA was used for each PCR reaction. If, according to the Qubit quantification, the DNA concentration was less than 2 ng mL<sup>-1</sup> or below the limit of detection, it was then used as is, and just 2 mL was added to the PCR reaction. PCR reactions were set up on ice, in a 1x reaction in 25 mL total volume. Final primer concentration was 0.5 mM and polymerase was Phusion Hot Start High-Fidelity Master Mix (Thermo Fisher Scientific Inc., Waltham, MA, USA). There were two cycles with different annealing temperatures, the first with an annealing temperature specific to the loci-specific region and the second set of cycles with an annealing temperature that also takes the MiSeq adapter sequence into account (Canesi and Rynearson 2016). PCR conditions used were initial denaturation for 30 seconds at 98 °C, 15 cycles of the following: denaturation for 10 seconds at 98 °C, annealing for 30 seconds at 64.1 °C, extension for 30 seconds at 72 °C, and 15 cycles with the same conditions except a higher annealing temperature of 72 °C, and then a final extension for 10 minutes at 72 °C, and a holding temperature of 10 °C until stored in the -20 °C freezer. PCR products were visualized on a 1% agarose gel before submission to the URI Genomics and Sequencing Center (Kingston, RI, USA) where library preparation and sequencing were performed on a 2x300 bp MiSeq run (Illumina, Inc., San Diego, CA, USA). There were 193 environmental samples were sequenced, along with two positive controls of *Pseudo-nitzschia* DNA from cultures and one negative control, for a total of 196 samples using two sets of MiSeq indices on the same sequencing plate. It was deemed appropriate to multiplex this plate as estimated read depth to recover *Pseudo-nitzschia* sequences was predicted to be lower than usual.

## Processing Description

A custom bioinformatics pipeline was utilized. CutAdapt (Martin 2011) was used to trim Illumina MiSeq adapters and primer sequences. Primer sequences were trimmed from both ends of sequences, with the reverse complement of the other primer trimmed the end of the sequences. If reads did not have the ITS1 primer sequence, they were discarded. Reads needed to be one base pair (bp) or longer to continue in the pipeline. Trimmed sequences were inputted into DADA2 (v. 1.16) to determine amplicon sequence variants (ASVs; Callahan et al. 2016). ASVs were retained at that level, with some potentially having as few differences as one bp to each other, for the subsequent analysis. ASVs were identified as *Pseudo-nitzschia* taxa using a curated database from NCBI sequences (Table S2 in Sterling et al. in prep) which used to design primers to assign taxonomy for ITS1 ASVs trimmed of the primer sequences using the scikit-learn naïve Bayes machine learning classifier (Pedregosa et al. 2011) at default settings in QIIME2 (Bolyen et al. 2019). The scikit-learn naïve Bayes machine learning classifier identified 97 ASVs as *Pseudo-nitzschia* at the species level. Three of these ASVs belonged to *P. subcurvata* from the positive control mock community and were removed from analysis. All of the 6,503 ASVs recovered from the 192 non-control samples from the sequencing effort were run through a megablast search using BLAST+ version 2.9 with the nucleotide (nt) database downloaded on October 4, 2020. There were 540 ASVs which had a known *Pseudo-nitzschia* taxa, including clones, vouchers, and environmental samples, as its top megablast hit. In addition to the 97 ASVs identified as a specific *Pseudo-nitzschia* species from the QIIME2 pipeline, there were 115 ASVs identified as a *Pseudo-nitzschia* taxa with greater than 75% query coverage were manually examined. It was determined by judgement call that the 11 ASVs which were identified as *P. pungens* PC50 were likely *Cylindrotheca* instead and the 85 ASVs which were closest related to *P. delicatissima* KJ22-0.2-69 environmental clone was most closely related to known *Nitzschia* isolate sequence from subsequent BLAST searches. This left 19 ASVs of interest, with 9 of them have >98% query coverage and >98% identity with known *Pseudo-nitzschia* sequences so were referred to as the specific *Pseudo-nitzschia* species and 10 ASVs were identified as the genus with identifiers of similar groups of ASVs to each other. These genus level ASVs have < 96% identity to existing sequences in the database. In total, there were 113 ASVs from the 192 samples that appeared to be of reliable *Pseudo-nitzschia* origin. Sample #AS424 had none of the 113 ASVs and was removed. Read counts were transformed into relative abundance out of total *Pseudo-nitzschia* taxa reads. If an ASV accounted for < 1% relative abundance in a sample, then it was considered "not present" or absent to avoid potentially spurious results. This removed 60 ASVs which only occurred in < 1% of reads in samples. The remaining 53 ASVs were used in the analysis in a presence/absence matrix to avoid potential problems from inflating read numbers with cell counts. This threshold retained 46 of the 97 scikit-learn classifier identified ASVs, and seven of the ASVs added by the megablast curation. Of the seven ASVs added from megablast

results, three ASVs were in a group together at the genus level, and around 95% identity with known *P. americana* sequences. The other megablast added ASVs were very closely related to *P. cuspidata* and *P. calliantha*.

#### **BCO-DMO Processing Notes:**

- data were submitted in file "DATA02\_ASV\_ID\_Sterling\_NBay.xlsx", Sheet 1 and extracted to csv.
- added conventional header with dataset name, PI name, version date
- renamed columns to conform with BCO-DMO naming conventions (removed hyphen)

[ [table of contents](#) | [back to top](#) ]

---

## **Related Publications**

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., ... Asnicar, F. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8), 852–857. doi:[10.1038/s41587-019-0209-9](https://doi.org/10.1038/s41587-019-0209-9)

*Methods*

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. doi:[10.1038/nmeth.3869](https://doi.org/10.1038/nmeth.3869)

*Methods*

Canesi, K., & Ryneerson, T. (2016). Temporal variation of *Skeletonema* community composition from a long-term time series in Narragansett Bay identified using high-throughput DNA sequencing. *Marine Ecology Progress Series*, 556, 1–16. doi:[10.3354/meps11843](https://doi.org/10.3354/meps11843)

*Methods*

Chappell, P., Armbrust, E., Barbeau, K., Bundy, R., Moffett, J., Vedamati, J., & Jenkins, B. (2019). Patterns of diatom diversity correlate with dissolved trace metal concentrations and longitudinal position in the northeast Pacific coastal-offshore transition zone. *Marine Ecology Progress Series*, 609, 69–86.

doi:[10.3354/meps12810](https://doi.org/10.3354/meps12810)

*Methods*

Hubbard, K. A., Rocap, G., & Armbrust, E. V. (2008). Inter- and Intraspecific Community Structure within the Diatom Genuspseudo-Nitzschia(Bacillariophyceae). *Journal of Phycology*, 44(3), 637–649.

doi:[10.1111/j.1529-8817.2008.00518.x](https://doi.org/10.1111/j.1529-8817.2008.00518.x)

*Methods*

Lelong, A., Hégaret, H., Soudant, P., & Bates, S. S. (2012). Pseudo-nitzschia (Bacillariophyceae) species, domoic acid and amnesic shellfish poisoning: revisiting previous paradigms. *Phycologia*, 51(2), 168–216.

doi:[10.2216/11-37.1](https://doi.org/10.2216/11-37.1)

*Methods*

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.

EMBnet.journal, 17(1), 10. doi:[10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200)

*Software*

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825–2830.

<https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>

*Software*

White, T. J., Bruns, T., Lee, S. J. W. T., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications*, 18(1), 315–322. <https://nature.berkeley.edu/brunslab/papers/white1990.pdf>

*Methods*

[ [table of contents](#) | [back to top](#) ]

---

## Parameters

| Parameter                            | Description   | Units    |
|--------------------------------------|---|----------|
| Sequence_ID                          | Sequence identifier associated with the sequence in NCBI's GenBank  | unitless |
| Sequence_of_ASV                      | The DNA sequence of the internal transcribed spacer (ITS) 1 region used to identify the Pseudo-nitzschia [NCBI:txid41953] species   | unitless |
| NCBI_GenBank_Accession_Number        | The unique accession number of each sequence in NCBI GenBank which ranges from MW447658-MW447770 covering all 113 sequences   | unitless |
| Pseudo_nitzschia_species             | The species of Pseudo-nitzschia that the ASV was determined to belong to. "Pseudo-nitzschia sp." indicates that no exact species could be determined and instead it was identified to the genus level.  | unitless |
| ASV_Number_on_Sterling_et_al_Figures | The arbitrary identification number assigned to the ASVs used in figures shown in the associated Sterling et al manuscript figures. nd = these ASVs were not used in figures  | unitless |
| ID_Method                            | The method used to identify the ASV as belonging to Pseudo-nitzschia sp. QIIME2 naiveBayes refers to the scikit-learn naïve Bayes machine learning classifier (Pedregosa et al 2011) at default settings in QIIME2 (Bolyen et al 2019) using the curated database of existing Pseudo-nitzschia sequences in NCBI in Table S2 of Sterling et al. Megablast refers to using BLAST+ version 2.9 with the nucleotide (nt) database and manually examining the identification of ASVs with query coverage > 75% to known Pseudo-nitzschia sequences.                     | unitless |
| Threshold_Pass                       | Whether or not the ASV passed the threshold of accounting for > 1% relative abundance in a sample. There were some ASVs which did not occur in > 1% relative abundance in any samples and were removed from the dataset altogether to avoid potentially spurious results. "Yes" refers to the ASVs which passed this relative abundance threshold and were included in subsequent data analysis and figures in Sterling et al and "no" refers to ASVs which never occurred at a threshold of >1% relative abundance in any of the samples across the whole dataset. | unitless |
| Notes                                | Any additional descriptions related to top hit of Pseudo-nitzschia species from the megablast search of the identification of the ASV. All ASVs with notes attached had   | unitless |

[ [table of contents](#) | [back to top](#) ]

## Instruments

|   |  |
|---|--|
| <b>Dataset-specific Instrument Name</b> | Illumina MiSeq Next Generation Sequencing (University of Rhode Island Genomics and Sequencing Center)  |
| <b>Generic Instrument Name</b>          | Automated DNA Sequencer  |
| <b>Generic Instrument Description</b>   | General term for a laboratory instrument used for deciphering the order of bases in a strand of DNA. Sanger sequencers detect fluorescence from different dyes that are used to identify the A, C, G, and T extension reactions. Contemporary or Pyrosequencer methods are based on detecting the activity of DNA polymerase (a DNA synthesizing enzyme) with another chemoluminescent enzyme. Essentially, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step. |

[ [table of contents](#) | [back to top](#) ]

---

## Deployments

### EN608

|                    |   |
|--------------------|---|
| <b>Website</b>     | <a href="https://www.bco-dmo.org/deployment/848016">https://www.bco-dmo.org/deployment/848016</a> |
| <b>Platform</b>    | R/V Endeavor  |
| <b>Start Date</b>  | 2018-01-31  |
| <b>End Date</b>    | 2018-02-06  |
| <b>Description</b> | C-AIM project   |

### EN617

|                   |   |
|-------------------|---|
| <b>Website</b>    | <a href="https://www.bco-dmo.org/deployment/848018">https://www.bco-dmo.org/deployment/848018</a> |
| <b>Platform</b>   | R/V Endeavor  |
| <b>Start Date</b> | 2018-07-20  |
| <b>End Date</b>   | 2018-07-25  |

### EN627

|                   |   |
|-------------------|---|
| <b>Website</b>    | <a href="https://www.bco-dmo.org/deployment/848056">https://www.bco-dmo.org/deployment/848056</a> |
| <b>Platform</b>   | R/V Endeavor  |
| <b>Start Date</b> | 2019-02-01  |
| <b>End Date</b>   | 2019-02-06  |

### EN644

|                   |   |
|-------------------|---|
| <b>Website</b>    | <a href="https://www.bco-dmo.org/deployment/848020">https://www.bco-dmo.org/deployment/848020</a> |
| <b>Platform</b>   | R/V Endeavor  |
| <b>Start Date</b> | 2019-08-20  |
| <b>End Date</b>   | 2019-08-25  |

[ [table of contents](#) | [back to top](#) ]

---

## Project Information

### **RII Track-1: Rhode Island Consortium for Coastal Ecology Assessment, Innovation, and Modeling (C-AIM)**

**Coverage:** Narragansett Bay, Rhode Island

NSF Award Abstract:

#### Non-technical Description

The University of Rhode Island (URI) will establish the Consortium for Coastal Ecology Assessment, Innovation, and Modeling (C-AIM) to coordinate research, education, and workforce development across Rhode Island (RI) in coastal marine science and ecology. C-AIM addresses fundamental research questions using observations, computational methods, and technology development applied to Narragansett Bay (NB), the largest estuary in New England and home to important ecosystem services including fisheries, recreation, and tourism. The research will improve understanding of the microorganisms in NB, develop new models to predict pollution and harmful algal bloom events in NB, build new sensors for nutrients and pollutants, and provide data and tools for stakeholders in the state. Observational capabilities will be coordinated in an open platform for researchers across RI; it will provide real-time physical, chemical, and biological observations ? including live streaming to mobile devices. C-AIM will also establish the RI STEAM (STEM + Art) Imaging Consortium to foster collaboration between artists, designers, engineers, and scientists. Research internships will be offered to undergraduate students throughout the state and seed funding for research projects will be competitively awarded to Primarily Undergraduate Institution partners.

#### Technical Description

C-AIM will employ observations and modeling to assess interactions between organisms and ecosystem function in NB and investigate ecological responses to environmental events, such as hypoxia and algal blooms. Observations of the circulation, biogeochemistry, and ecosystem will be made using existing and new instrument platforms. The Bay Observatory ? a network of observational platforms around NB - will be networked to trigger enhanced water sampling and sensing during specific environmental events, such as hypoxic conditions or phytoplankton blooms. Biogeochemical, ecological, and coastal circulation models will be integrated and coupled to focus on eutrophication and pollutant loading. Data and models will be integrated on multiple scales, from individual organisms and trophic interactions to food-web responses, and from turbulence to the regional ocean circulation. New sensing technologies for nutrients and pollutants will be developed, including affordable, micro-fluidic (Lab-on-a-Chip) devices with antifouling capabilities. The results will be synthesized and communicated to stakeholders.

[ [table of contents](#) | [back to top](#) ]

---

## Funding

| <b>Funding Source</b>                                    | <b>Award</b>                   |
|--|--------------------------------|
| <a href="#">NSF Division of Ocean Sciences (NSF OCE)</a> | <a href="#">OCE-1655686</a>    |
| <a href="#">NSF Division of Ocean Sciences (NSF OCE)</a> | <a href="#">OIA-1655221</a>    |
| National Oceanic and Atmospheric Administration (NOAA)   | <a href="#">NA18OAR4170094</a> |
| National Oceanic and Atmospheric Administration (NOAA)   | <a href="#">NA14OAR4170082</a> |

[ [table of contents](#) | [back to top](#) ]