# Fragment analysis files from microsatellite analysis of samples of the basket cockle (Clinocardium nuttallii) from multiple sites in Washington State between Mar of 2019 and Jan of 2020

**Website**: https://www.bco-dmo.org/dataset/940678
**Data Type**: Other Field Results
**Version**: 1
**Version Date**: 2024-10-21

**Project**
» Quantifying and modeling the transmission dynamics of bivalve transmissible neoplasia (Transmission of BTN)


| Contributors | Affiliation | Role |
|---|---|---|
| Metzger, Michael J. | Pacific Northwest Research Institute (PNRI) | Principal Investigator |
| Crim, Ryan | Puget Sound Restoration Fund | Co-Principal Investigator |
| Unsell, Elizabeth | Suquamish Tribe | Co-Principal Investigator |
| Abbott, Cathryn | Fisheries and Oceans Canada, Pacific Region (DFO MPO) | Scientist |
| Dimond, James | Western Washington University (WWU) | Scientist |
| Gurney-Smith, Helen | Fisheries and Oceans Canada, Pacific Region (DFO MPO) | Scientist |
| Little Wing Sigo, Robin | Suquamish Tribe | Scientist |
| Smith, Peter | Pacific Northwest Research Institute (PNRI) | Scientist |
| Supernault, Janine | Fisheries and Oceans Canada, Pacific Region (DFO MPO) | Scientist |
| Vandepas, Lauren | University of Miami | Scientist |
| Weinandt, Sydney | Pacific Northwest Research Institute (PNRI) | Scientist |
| Withler, Ruth | Fisheries and Oceans Canada, Pacific Region (DFO MPO) | Scientist |
| Child, Zachary | Pacific Northwest Research Institute (PNRI) | Technician |
| Garrett, Fiona | Pacific Northwest Research Institute (PNRI) | Technician |
| Giersch, Rachael | Pacific Northwest Research Institute (PNRI) | Technician |
| Sevigny, Jordana | Pacific Northwest Research Institute (PNRI) | Technician |
| Yonemitsu, Marisa | Pacific Northwest Research Institute (PNRI) | Technician |
| York, Amber D. | Woods Hole Oceanographic Institution (WHOI BCO-DMO) | BCO-DMO Data Manager |


**Abstract**
These files support the publication "Multiple lineages of transmissible neoplasia in the basket cockle (C. nuttallii) with repeated horizontal transfer of mitochondrial DNA," which has been submitted to bioRxiv (doi:10.1101/2023.10.11.561945). Details of locations and dates of cockle collection, specific primer sequences, methods for design of the primers, and analysis of results can be found there. These files are fragment analysis files from microsatellite analysis of samples of the basket cockle (Clinocardium nuttallii) from multiple sites in Washington State. For each cockle DNA was extrcted from a solid tissue sample (T) and a hemolymph sample (H). These were used to identify bivalve transmissible neoplasia in several cockle samples in these populaitons (together with other genetic data from other nuclear and mitochondrial markers). Microsatellites were amplified from DNA extracted from both hemocyte and tissue using Taq polymerase (Genesee Scientific) for a select number of individuals: all possible neoplastic animals based on positive qPCR screen (n=26) and one randomly selected, non-neoplastic animal from each collection (n=12). Microsatellites were amplified using primers for 8 polymorphic loci: Cnu48, Cnu55, Cnu58, Cnu63, Cnu68, Cnu72, Cnu78, and Cnu81. Allele sizes were identified by fragment analysis using fluorescent primers (6-FAM, PET, NED, and VIC) and a 3730xl Genetic Analyzer with the LIZ-500 size standard (Applied Biosystems, operated by Genewiz). For each sample, two reactions were run: "1" with (Cnu48, blue; Cnu55, yellow; Cnu58, red; Cnu63, green) and "2"

with (Cnu68, blue; Cnu72, yellow; Cnu79, red; Cnu81, green).

---

## Table of Contents



---

## Coverage

**Location**: Washington State, USA
**Spatial Extent**: **N**:48.983317 **E**:-122.5539 **S**:46.68 **W**:-124.62356
**Temporal Extent**: 2019-03-18 - 2020-01-09

### Methods & Sampling

Microsatellites were amplified from DNA extracted from both hemocyte and tissue using Taq polymerase (Genesee Scientific) for a select number of backet cockles (Clinocardium nuttallii, urn:lsid:marinespecies.org:taxname:381980) colelcted from multipl intertidal locations in Washington State, USA. All possible neoplastic animals based on positive qPCR screen (n=26) and one randomly selected, non-neoplastic animal from each collection (n=12) were selected for microsatellite amplification and fragment analysis. Microsatellites were amplified using primers for 8 polymorphic loci: Cnu48, Cnu55, Cnu58, Cnu63, Cnu68, Cnu72, Cnu78, and Cnu81.

Instrument Description: Allele sizes were identified by fragment analysis using fluorescent primers (6-FAM, PET, NED, and VIC) and a 3730xl Genetic Analyzer with the LIZ-500 size standard.

### Data Processing Description

Data uploaded are raw data files, before processing.

### BCO-DMO Processing Description

* All .fsa files submitted to BCO-DMO under folder "All Cnu microsatellites combined/" were bundled into "All_Cnu_microsatellites_combined.zip."
* Site information and collection metadata was extracted from a tab delimited table provided in metadata and imported into the BCO-DMO data system as a table. A location's lat,lon was revised based on the data submitter's input ("48 58,999", "-122 47,608" ->"48.98332", "-122.79347").
* ISO datetime with timezone added in UTC time zone (converted using the local date and time provided by submiter in PST/PDT).
* A file inventory was created including md5sum. The site_code was extracted as a dedicated column using the file prefix. The site_code was used to join the metadata table after verifying site_code was a unique key in the metadata table. File inventory and combined collection metadata attached to teh dataset as "940678_v1_cnu-microsatellites.csv"
* The Sample/Collection table alone was also attached as a supplemental file (row per site/collection).

* Column names adjusted to conform to BCO-DMO naming conventions designed to support broad re-use by a variety of research tools and scripting languages. [Only numbers, letters, and underscores.  Can not start with a number]

* Rows with site code "SBg" were dropped at data submitter's requrest. (see Problems/Issues section)

**Problem Description**

Several samples were rerun with more DNA in cases with high background. These files are marked with a suffix after an underscore (e.g. "SB19H-2_r" is the second run of the same sample as "SB19H-2").

All samples from site "SBg"("Sequim Bay geoduck tubes, WA") were excluded from this dataset (and results publication) due to a possible sample mix-up in a small number of the samples analyzed.

[ table of contents | back to top ]

## Data Files

| File |
|---|
| **Fragment analysis files (.fsa) from Clinocardium nuttallii microsatellites**<br>filename: All_Cnu_microsatellites_combined.zip (ZIP Archive (ZIP), 26.36 MB)<br>MD5:37d476f7d0fce041d146fd7c9bd62642<br><br>Fragment analysis files (.fsa format).  These files comprise a set of microsatellite loci that have been amplified by different primers. For each sample, there are 8 different PCR reactions, each with a labeled primer. They are multipexed in sets of 4, since there are 4 available dyes. So there are 2 files for each sample. These data can be analyzed to determine the microsatellite allele sizes.<br><br>See site information and collection metadata for these files contained in "940678_v1_cnu-microsatellites.csv"<br><br>FSA file format is used by proprietary software (Peak Scanner by Thermo Fisher and GeneMapper).  They can also be used by open source tools (e.g. biopython).<br><br>FSA File Structure:<br>The FSA file contains multiple data blocks, each with specific information about the sequencing run. These include metadata (run conditions, sample name) and raw data (electropherogram peaks).<br>Data Blocks:<br>The data in FSA files is organized in tagged data blocks (each block contains a structured data type (e.g. an array, a sequence, etc). |

[ table of contents | back to top ]

## Supplemental Files

## File

### Fragment analysis file inventory and collection metadata

filename: 940678_v1_cnu-microsatellites.csv

(Comma Separated Values (.csv), 28.00 KB)
MD5:458c0c945443950eaf2e8412517fab87

File metadata table including collection information for all .fsa files included in "All_Cnu_microsatellites_combined.zip."

Columns (Parameters):

column_name,column_description, units, data_type,format

filename,Fragment analysis filename (.fsa),units,String,

filesize_bytes,filesize,bytes,Integer,

md5sum,checksum (md5 hash) which can be used to verify intergrity of transferred files.,unitless,String,

site_code,Site Code,unitless,String,

Location,"Site location description (e.g. ""Sequim Bay geoduck tubes, WA"")",unitless,String,

Beach_Name,"Beach name (e.g. ""Front Beach"")",unitless,String,

Collection_Date,Collection date (local time zone PST/PDT),unitless,Date,%m/%d/%Y

Tribe_Agency,Tribal Agency,unitless,String,

GPS_Lat,Site latitude for collection,decimal degrees,Float,

GPS_Lon,Site longitude for collection,decimal degrees,Float,

Time,Collection time (local time zone PST/PDT),unitless,Time,%H:%M

Tidal_Height,Tidal height at time of collection,feet (ft),Float,

Collection_DateTime_UTC,Datetime with timezone for collection (UTC),units,Datetime,%Y-%m-%dT%H:%MZ

### Sample Log (and Site List)

filename: sample_log.csv

(Comma Separated Values (.csv), 1.32 KB)
MD5:19322b3cc7ead4348774d3f1a4367c91

This table contains a sample log for cockles collected from multiple intertidal locations in Washington State, USA. It includes a row per site along with the site_code used in the .fsa filenames.

Columns (Parameters):

column_name,column_description, units, data_type,format

Location,"Site location description (e.g. ""Sequim Bay geoduck tubes, WA"")",unitless,String,

Beach_Name,"Beach name (e.g. ""Front Beach"")",unitless,String,

Code,Site Code,unitless,String,%m/%d/%Y

Collection_Date,Collection date (local time zone PST/PDT),unitless,Date,%m/%d/%Y

Tribe_Agency,Tribal Agency,unitless,String,

GPS_Lat,Site latitude for collection,decimal degrees,Float,

GPS_Lon,Site longitude for collection,decimal degrees,Float,

Time,Collection time (local time zone PST/PDT),unitless,Time,%H:%M

Tidal_Height,Tidal height at time of collection,feet (ft),Float,

Collection_DateTime_UTC,Datetime with timezone for collection (UTC),units,Datetime,%Y-%m-%dT%H:%MZ

[ table of contents | back to top ]

## Related Publications

Yonemitsu, M. A., Sevigny, J. K., Vandepas, L. E., Dimond, J. L., Giersch, R. M., Gurney-Smith, H. J., Abbott, C. L., Supernault, J., Withler, R., Smith, P. D., Weinandt, S. A., Garrett, F. E. S., Sigo, R. L. W., Unsell, E., Crim, R. N., & Metzger, M. J. (2023). Multiple lineages of transmissible neoplasia in the basket cockle (Clinocardium nuttallii) with repeated horizontal transfer of mitochondrial DNA. https://doi.org/10.1101/2023.10.11.561945 *Results*

[ table of contents | back to top ]

## Parameters

| Parameter | Description | Units |
|---|---|---|
| filename | Filename of the .fsa file (included in All_Cnu_microsatellites_combined.zip) | unitless |
| filesize_bytes | filesize in bytes | bytes |
| md5sum | checksum (md5 hash) which can be used to verify intergrity of transferred files. | unitless |
| site_code | Site Code | unitless |
| Location | Site location description | unitless |
| Beach_Name | Beach name (e.g. "Front Beach") | unitless |
| Collection_Date | Collection date (local time zone PST/PDT) | unitless |
| Tribe_Agency | Tribal Agency | unitless |
| GPS_Lat | Site latitude for collection | decimal degrees |
| GPS_Lon | Site longitude for collection | decimal degrees |
| Time | Collection time (local time zone PST/PDT) | unitless |
| Tidal_Height | Tidal height at time of collection | feet (ft) |
| Collection_DateTime_UTC | Datetime with timezone for collection (UTC) | unitless |

[ table of contents | back to top ]

## Instruments

| | |
|---|---|
| **Dataset-specific Instrument Name** | 3730xl Genetic Analyzer with the LIZ-500 size standard |
| **Generic Instrument Name** | Automated DNA Sequencer |
| **Generic Instrument Description** | 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 ]

## Project Information

**Quantifying and modeling the transmission dynamics of bivalve transmissible neoplasia (Transmission of BTN)**

**Coverage**: East and West coasts of USA

*NSF Award Abstract:*
Cancer is not normally thought of as an infectious disease. However, several transmissible cancers have recently been found in the wild, in which the cancer cells themselves jump from animal to animal as an infectious agent, causing significant mortality on land and in the marine environment. Marine bivalves appear to be particularly susceptible. At least nine lineages of lethal transmissible cancer have been identified in eight bivalve species worldwide since they were first recognized as an infectious cancer by members of this team less than a decade ago. It is known that whole cancer cells transfer from one animal to another, but it is unclear how this infectious disease spreads at the individual level, within a single population, or between populations in the environment. The interdisciplinary team is combining sensitive field surveys of disease prevalence, laboratory inoculation, and in vitro experiments together with quantitative modeling to understand how this unique class of infectious disease spreads in nature. The team will continue to communicate the results of this project through scientific publications and meetings with commercial aquaculture and local Native American communities, including research partners in multiple Coast Salish Tribes. Understanding the disease transmission principles may help develop strategies to control this disease, which would directly help these communities. The team members are also training undergraduate students during summer research experiences at Pacific Northwest Research Institute, Western Washington University, and Bigelow Laboratory for Ocean Sciences.

To understand the basic principles of the spread of bivalve transmissible cancer, the team studies two separate lineages in geographically separated species: soft-shell clams (Mya arenaria) on the Atlantic Coast of North America, the first bivalve transmissible cancer identified; and basket cockles (Clinocardium nuttallii), a species on the Pacific Coast of North America, in which the team has just recently identified bivalve transmissible cancer. The team is developing two quantitative models, one for the spread of disease within a population over time, and a second to model the spread of cancer lineages between different populations. They are testing these models with regular disease prevalence data from wild populations from multiple sites. Laboratory work on disease progression and transmission supports development and refinement of these models by providing critical parameter values and testing whether environmental variables (such as temperature) or genetic variables (such as the relatedness of cancer and host) affect the susceptibility and timing of disease progression. This project aims to develop a quantitative understanding of disease dynamics in soft-shell clams and basket cockles. Ultimately, it will provide general principles that underlie the spread of this recently discovered class of infectious disease.

This project was funded by the Division of Environmental Biology and the Division of Ocean Sciences.

This award reflects NSF's statutory mission and has been deemed worthy of support through evaluation using the Foundation's intellectual merit and broader impacts review criteria.

[ table of contents | back to top ]

## Funding


| Funding Source | Award |
| --- | --- |
| NSF Division of Ocean Sciences (NSF OCE) | OCE-2208081 |


[ table of contents | back to top ]