

*Please provide the following information, and submit to the NOAA DM Plan Repository.*

### **Reference to Master DM Plan (if applicable)**

*As stated in Section IV, Requirement 1.3, DM Plans may be hierarchical. If this DM Plan inherits provisions from a higher-level DM Plan already submitted to the Repository, then this more-specific Plan only needs to provide information that differs from what was provided in the Master DM Plan.*

URL of higher-level DM Plan (if any) as submitted to DM Plan Repository:

## **1. General Description of Data to be Managed**

### **1.1. Name of the Data, data collection Project, or data-producing Program:**

Vibrio population structure - Genetic and population structure analysis of clinical and environmental *Vibrio parahaemolyticus* strains

### **1.2. Summary description of the data:**

*Vibrio parahaemolyticus* (Vp) is a marine bacterium capable of causing severe gastroenteritis in humans, usually through the consumption of raw shellfish. Before 1995, Vp-vibriosis was sporadic world-wide and caused by a relatively heterogeneous population of the bacterium. Since then, outbreaks have become more epidemic, with foci of infections traced to seafood harvested from single or geographically-linked sites. While initial outbreaks in Asia (and later in South America and the U.S. Gulf Coast region) have been attributed to a single serotypically-related pandemic clonal complex, other serotypes have been implicated in distinct geographical areas, including the Pacific Northwest and Alaska in the U.S.

Current risk assessment models are based on the presence of the virulence-associated genes *tdh* and *trh*, yet illnesses have been attributed to *tdh*- and/or *trh*- isolates. Previous phylogenetic studies have shown that Vp, like most *Vibrio* spp., is a genetically diverse species, and as yet there has been no definitive conclusion as to what genes are essential for virulence. Using phenotypic, genetic, and genomic comparison methods such as Multi-Locus Sequence Typing (MLST), we are examining the hypothesis that a set of highly-virulent clones of Vp with increased pathogenic potential have recently emerged in the PNW, and determining whether the emergence can be correlated with specific environmental parameters. MLST and other genotyping analyses of clinical and environmental Vp isolates from PNW sources demonstrate the extensive patterns of diversity as seen elsewhere. However, the majority of PNW strains obtained from human infections form a distinct clonal complex separate from most environmental isolates. Interestingly, many environmental isolates obtained from PNW sources are phylogenetically related to the pandemic clonal complex, but this group has not been associated with clinical infections in the region.

Genome sequences.

### **1.3. Is this a one-time data collection, or an ongoing series of measurements?**

One-time data collection

**1.4. Actual or planned temporal coverage of the data:**

2008-10-01 to 2012-10-05

**1.5. Actual or planned geographic coverage of the data:**

W: -122.3062, E: -122.3062, N: 47.6449, S: 47.6449

NWFSC Montlake: NWFSC Montlake lab, Seattle

**1.6. Type(s) of data:**

*(e.g., digital numeric data, imagery, photographs, video, audio, database, tabular data, etc.)*

Other

**1.7. Data collection method(s):**

*(e.g., satellite, airplane, unmanned aerial system, radar, weather station, moored buoy, research vessel, autonomous underwater vehicle, animal tagging, manual surveys, enforcement activities, numerical model, etc.)*

**1.8. If data are from a NOAA Observing System of Record, indicate name of system:**

**1.8.1. If data are from another observing system, please specify:**

**2. Point of Contact for this Data Management Plan (author or maintainer)**

**2.1. Name:**

Northwest Fisheries Science Center (NWFSC)

**2.2. Title:**

Metadata Contact

**2.3. Affiliation or facility:**

Northwest Fisheries Science Center (NWFSC)

**2.4. E-mail address:**

nmfs.nwfsc.metadata@noaa.gov

**2.5. Phone number:**

206-860-3200

**3. Responsible Party for Data Management**

*Program Managers, or their designee, shall be responsible for assuring the proper management of the data produced by their Program. Please indicate the responsible party below.*

**3.1. Name:**

Mark S Strom

**3.2. Title:**

Data Steward

#### 4. Resources

*Programs must identify resources within their own budget for managing the data they produce.*

**4.1. Have resources for management of these data been identified?**

No

**4.2. Approximate percentage of the budget for these data devoted to data management (specify percentage or "unknown"):**

25

#### 5. Data Lineage and Quality

*NOAA has issued Information Quality Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information which it disseminates.*

**5.1. Processing workflow of the data from collection or acquisition to making it publicly accessible**

*(describe or provide URL of description):*

Lineage Statement:

variety of bioinformatics-based analyses

**5.1.1. If data at different stages of the workflow, or products derived from these data, are subject to a separate data management plan, provide reference to other plan:**

**5.2. Quality control procedures employed (describe or provide URL of description):**

Sequence reads were filtered through SOLiD Accuracy Enhancement Tool (SAET) and PCR duplicated reads were removed using the fastq\_nodup tool from the SEASAR pipeline . Genomes were assembled in color-space using the CLC Assembly Cell version 4.1.

#### 6. Data Documentation

*The EDMC Data Documentation Procedural Directive requires that NOAA data be well documented, specifies the use of ISO 19115 and related standards for documentation of new data, and provides links to resources and tools for metadata creation and validation.*

**6.1. Does metadata comply with EDMC Data Documentation directive?**

No

**6.1.1. If metadata are non-existent or non-compliant, please explain:**

Missing/invalid information:

- 1.7. Data collection method(s)

**6.2. Name of organization or facility providing metadata hosting:**

NMFS Office of Science and Technology

**6.2.1. If service is needed for metadata hosting, please indicate:**

**6.3. URL of metadata folder or data catalog, if known:**

<https://www.fisheries.noaa.gov/inport/item/17957>

**6.4. Process for producing and maintaining metadata**

*(describe or provide URL of description):*

Metadata produced and maintained in accordance with the NOAA Data Documentation Procedural Directive:

[https://nosc.noaa.gov/EDMC/DAARWG/docs/EDMC\\_PD-Data\\_Documentation\\_v1.pdf](https://nosc.noaa.gov/EDMC/DAARWG/docs/EDMC_PD-Data_Documentation_v1.pdf)

**7. Data Access**

*NAO 212-15 states that access to environmental data may only be restricted when distribution is explicitly limited by law, regulation, policy (such as those applicable to personally identifiable information or protected critical infrastructure information or proprietary trade information) or by security requirements. The EDMC Data Access Procedural Directive contains specific guidance, recommends the use of open-standard, interoperable, non-proprietary web services, provides information about resources and tools to enable data access, and includes a Waiver to be submitted to justify any approach other than full, unrestricted public access.*

**7.1. Do these data comply with the Data Access directive?**

Yes

**7.1.1. If the data are not to be made available to the public at all, or with limitations, has a Waiver (Appendix A of Data Access directive) been filed?**

**7.1.2. If there are limitations to public data access, describe how data are protected from unauthorized access or disclosure:**

**7.2. Name of organization of facility providing data access:**

Northwest Fisheries Science Center (NWFS)

**7.2.1. If data hosting service is needed, please indicate:**

No

**7.2.2. URL of data access service, if known:**

[https://www.webapps.nwfsc.noaa.gov/parr/Vibrio\\_Population\\_Structure\\_Supplementary\\_material\\_4](https://www.webapps.nwfsc.noaa.gov/parr/Vibrio_Population_Structure_Supplementary_material_4)

**7.3. Data access methods or services offered:**

Genbank, <http://www.ncbi.nlm.nih.gov/genbank/>, Accession numbers: AONA00000000

AOOV00000000

AOOW00000000

AOOX00000000

AOOY00000000

AOOZ00000000

AOPA00000000  
AOPB00000000  
AOPC00000000  
AOPD00000000  
AOPE00000000  
AOPF00000000  
AOPG00000000  
AOOU00000000  
AOPH00000000  
AOPJ00000000  
AOPK00000000  
AOPM00000000  
AOPN00000000  
AOPP00000000  
AOPQ00000000  
AOPR00000000  
AOPS00000000  
AOPT00000000  
AOPU00000000  
AOPV00000000  
AOPW00000000  
AOPX00000000  
AOPY00000000  
AOPZ00000000  
AOPA00000000

**7.4. Approximate delay between data collection and dissemination:**

0 days

**7.4.1. If delay is longer than latency of automated processing, indicate under what authority data access is delayed:**

No Delay

**8. Data Preservation and Protection**

*The NOAA Procedure for Scientific Records Appraisal and Archive Approval describes how to identify, appraise and decide what scientific records are to be preserved in a NOAA archive.*

**8.1. Actual or planned long-term data archive location:**

*(Specify NCEI-MD, NCEI-CO, NCEI-NC, NCEI-MS, World Data Center (WDC) facility, Other, To Be Determined, Unable to Archive, or No Archiving Intended)*

OTHER

**8.1.1. If World Data Center or Other, specify:**

**8.1.2. If To Be Determined, Unable to Archive or No Archiving Intended, explain:**

**8.2. Data storage facility prior to being sent to an archive facility (if any):**

Northwest Fisheries Science Center - Seattle, WA

**8.3. Approximate delay between data collection and submission to an archive facility:**

365

**8.4. How will the data be protected from accidental or malicious modification or deletion prior to receipt by the archive?**

*Discuss data back-up, disaster recovery/contingency planning, and off-site data storage relevant*

*to the data collection*

Data is maintained in a US Government data repository for genetic sequence information (Genbank)

**9. Additional Line Office or Staff Office Questions**

*Line and Staff Offices may extend this template by inserting additional questions in this section.*