

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D1Z60Q>

Title: Genomic Standards Consortium (GSC) Island Sampling Day: Moorea Reef to Ridges Genomic Transect

Creator: Erin Robinson - **ORCID:** [0000-0001-9998-0114](https://orcid.org/0000-0001-9998-0114)

Affiliation: Tetiaroa Society

Principal Investigator: Neil Davies , Ramona Walls, Serge Planes, Chris Meyer, Lynn Schriml

Contributor: Scott Tighe, Pier Luigi Buttigieg, Raïssa Meyer

Funder: Genomic Standards Consortium

Template: Gump South Pacific Research Station

Project abstract:

Here we describe a project that supports the mission of the Genomic Standards Consortium (GSC) and contributes to the “Ocean Biomolecular Observations Network” (OBON) and the “Ocean Best Practice System” (OBPS - Omics Task Team), flagship programs of the UN Ocean Decade of Ocean Science for Sustainable Development. The project serves as a test case for two related infrastructure projects that aim to improve standards, policies, and best practices in sample collection: Sampling Nature (SN) Research Coordination Network and the Internet of Samples (iSamples).

The project’s primary goal (first paper) is methodological: (i) to demonstrate the value of genomic standards for both genomic research and secondary interdisciplinary research, (ii) to highlight the importance of data curation workflows that follow FAIR+CARE principles maximizing the potential for appropriate reuse of data and material samples, (iii) to illustrate the challenges that remain in implementing those standards - including ethical, legal, and social aspects, and (iv) to demonstrate some of the tools, infrastructures, and best-practices available for overcoming these challenges. This paper will include all the elements of a data paper (primary publication of the initial data collected), but with broader discussion around it as a use case for metadata standards and best practices in multi-omic sampling.

The project's secondary goal (second paper) is to address research questions including : (1) To what extent can one day of environmental sampling, and subsequent multi-omic analyses, characterize the ecological state of an island (coupled marine-terrestrial ecosystem)? (2) To what extent might such surveys serve as baselines for futuromic study (future analysis of biomolecules from the material samples archived), and (3) if repeated, could these surveys help address long-term changes in genomic biodiversity? (4) How does marine and terrestrial eDNA change across a transect? Moorea is one of the most well-described biotas in the world and has one of the longest and most intensive time-series of ecological data, particularly for coral reef ecosystems. It thus serves as a powerful calibration site for such a study.

The project's broader impact goal is public outreach – if possible engage local citizen scientists/schools, ideally through Fare Natura (ecomuseum at entrance to the Opunohu Valley. (Also perhaps local Associations that might be interested, e.g., Te Pu Atitia). Sampling Nature will produce an outreach videos based on this Island Sampling Day.

Start date: 03-26-2022

Last modified: 05-10-2023

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Genomic Standards Consortium (GSC) Island Sampling Day: Moorea Reef to Ridges Genomic Transect

Methodology

How will data be collected or produced?

We plan to sample a transect along an entire rivershed from the open ocean through the lagoon and bay up into the rivershed to the mountain ridges. We plan 20 stations across the ~8 km transect from -50m deep to 600 m. There will be 15 terrestrial stations from Ta'ahiamanu Beach (sea level, near the mouth of the bay) to the top of the main branch of the Opunohu River (600) and 5 marine stations spaced evenly from the mouth of the Opunohu River to the reef. All stations will sample aquatic environments in the water column (surface level for terrestrial stations and xx m below surface for marine stations) and, where accessible, the sediment below. For terrestrial stations, we will also sample soil (3 m from the bank of flowing streams or from the center in the case of seasonally dry streams). At terrestrial sites, we will collect 2 replicate 1 L water samples, 2 replicate 50 mL sediment samples, and 2 replicate 50 mL soil samples. Water samples will be filtered the same day, and only filters will be retained. In marine sites, 20 L of water will be filtered in situ and 2 replicate 50 mL sediment samples will be collected where possible. Field sampling metadata will be recorded using the GEOME platform.

We will target two different environmental sample types:

- Soil and Sediment protocols
 - [NEON protocol](#)
- Water Column protocols:
 - CRIOBE
 - 20L marine
 - [OSD protocol \(handbook\)](#)
 - [OSD 22](#)

Field Sampling

Water column

Two sets of sample water will be collected, 20L of water for standard processing via sterivac and 1L for Innovaprep filtering. The 1L water samples will be collected from the stream starting at Belvedere Point on [Mount Tohivea](#) to the bay and in Opunohu Bay, collected aseptically using a 1 DNA-free bottles, sealed, and transported to the the Gump research station for sample concentration using sterivac and InnovaPrep hollow fiber concentrator pipet (CP). Final elution of sample from InnovaPrep will be performed using 1ml of the CP elution buffer. Five-20L water samples will be collected by CRIOBE researchers in Opunohu Bay, every 500 meters. Concentrated samples will be split to two 1.5ml screw capped tubes and stored at -20C until QC and DNA extraction. Negative controls will be processed simultaneously side by side with sample processing.

Positive DNA controls will be spiked into a PBS prior to DNA extraction and library prep.

- Equipment and Supplies needed
 - Backpacks
 - 1 liter DNA free bottles or whirl-paks
 - Gloves
 - Innovaprep CP
 - 0.2 uM Innovaprep CP hollow fiber concentrator tips
 - CP elution buffer
 - 1.5 mL Screw capped tubes
 - Sharpies

Opunohu Bay soil, water and sediment samples will be collected in a 50 mL tube will be collected along the shoreline, at each of five sites along the same transect.

Aquatic sediment samples will be collected using a downstream to upstream orientation using 50 mL centrifuge tubes. Using new gloves for each sampling site, a 50 mL tube will be submerged in the middle of the stream below the area the water sample was collected at an inverted orientation until the opening of the tube penetrates the sediment to a depth of 10 cm. Using a downstream to upstream direction, the 50 mL tube is pushed as deep into the sediment as possible and scooped in an upstream motion to collect below base sediment. The tube is capped immediately.

Soil samples will be collected within 1m of the stream. Any leaf litter will be swept away by hand. The top 5-10 cm of soil will be removed.

Soil Samples will be collected using 70% ethanol sterilized metal scoops by aseptically scooping bank soil into 50 mL sterile DNA-free conical centrifuge tubes. Tubes will be capped and transported to the research station for -20 C storage until DNA extraction.

Equipment and Supplies :

- 80 - 50mL Falcon centrifuge tubes
- Gloves
- Flame sterilize metal spoons

- Sharpie
- 70% ethanol wipes
- Metal spoons
- Bags to carry supplies
- pH paper-Scott
- 50 1.5 ml tubes -Scott

Sample Processing on Moorea

The primary goal of sample processing onsite in Moorea is to preserve the materials for shipping to participating repositories (for long-term archiving) and participating labs (for immediate analyses).

Water columns:

Sterivex Innovaprep

Water samples concentrated by Innovaprep from up to 1L to 1ml for DNA extraction for a total of 20- 40 water samples. Samples will be eluted using 1ml CP elution buffer and transferred to two 1.5 ml screw capped tubes using a P1000.

- Equipment and supplies needed
 - Innovaprep CP
 - Innovaprep CP 0.2uM concentrating pipets
 - CP elution Buffer
 - Pipette
 - 1.5 ml screw capped tubes

Soil:

Shipping / Export

Material samples will be shipped in [insert protocols]

Omic Analyses

DNA extraction and Quality Control will be performed from up to 500ul of sample (either concentrated further by centrifugation at 5000xg 10 min or direct extraction) by adjusting to a final concentration of 1X PBS and pretreating with 5ul of metapolyzyme at 35 C for 1-4hr and extracted using the manufacturers

recommended procedures for the Qiagen Powersoil DNA extraction kit. QA/QC: Quantitation will be performed by Qubit spectrofluometer, purity by Nanodrop Spectrophotometer, and DNA size assessment using Agilent Bioanalyzer.

- Metapolyzyme
- Qiagen DNA Power Soil kit
- Qubit
- Nanodrop
- Agilent Bioanalyzer 2100
- Pipets

Microbiome and Metagenomic Library synthesis:

Library Synthesis for Illumina, Genapsys, PacBio will be accomplished using the recommended library kits described by the sequencing platform. For Illumina, either the Lucigen NXSeq, Nextera, NEB, or Qiagen will be used. Special library synthesis will be selected for others as indicated. Library Synthesis for Oxford Nanopore will be accomplished using either RBK0004 (rapid transposase barcoding), NBD114 (Ligation barcoding), or RPB004 (Rapid PCR barcoding) depending on available input DNA.

Next Generation Sequencing will be performed on an Illumina MiSeq (or equivalent) as well Oxford Nanopore, (MK1C, GridIonX5, PromethIon) PacBio, and/or Genapsys depending on interest.

Nutrient Analyses

Water column

Access, Data Sharing and Reuse

Will you require an embargo period prior to making your data available? If requested, an embargo period may be granted for up to [1 year] after the end date of the Project as specified in its Data Management Plan.

- No

Do you agree to share all data under the CC-0 license?

- Yes

Will your project include the collection of material samples? For example, archeological, geochemical (geosamples), and biological (biosamples) materials.

- Yes

Samples will include biologic and geosamples

Please describe standards you will utilize to register sampling events, apply unique identifiers, implement relevant metadata standards, and track derived material samples, data, and outputs.

Metadata describing the sampling site, collected sample and environmental parameters will be recorded on prepared metadata recording sheets, provided to each sampling group. The metadata terms, from the Genomic Standards Consortium's (www.genesc.org) MIxS minimal information metadata standard (<https://www.ncbi.nlm.nih.gov/biosample/docs/packages/MIMARKS.survey.soil.5.0/>). Chemical analysis of soil samples and DNA extraction metadata will be recorded in the metadata recording sheets.

What are the further intended and/or foreseeable research uses for the completed dataset(s)?

The completed datasets of soil, sediment and concentrated water samples will be stored for the analysis of organismal environmental DNA (eDNA) and microbial DNA found in the terrestrial and aquatic environments, for identifying the predominant species and assessing the variability of organisms along the transect and between the soil, sediment and water environments.

State any expected difficulties in data sharing, along with causes and possible measures to overcome these difficulties.

Question not answered.

Documentation and Metadata

What documentation and metadata will accompany the data?

The recorded metadata will include geographic location, lat/long, collection date/time, the type of site (water - river, water - ocean; sediment - river; sediment - ocean, soil - by river), sampling depth in the water, pH, air temperature, humidity, soil, sediment, and water temperature.

Ethics and Intellectual Property

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues? Demonstrate that you have sought advice on and addressed all copyright and rights management issues that apply to the resource.

Question not answered.

How will you handle sensitive data? Make explicit mention of consent, confidentiality, anonymization, and other ethical considerations, where appropriate.

Question not answered.

Are any restrictions on data sharing required – for example, to safeguard research participants or to gain appropriate intellectual property protection?

- No

Describe restrictions on data sharing required due to privacy or IP protection.

Question not answered.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

All data is of long-term value and will be retained, shared and preserved.

What is the long-term preservation plan for the dataset?

Genomic sequence data will be preserved in INSDC

Planned Research Outputs

Dataset - "Genetic sequences from soil, sediment and concentrated water samples "

Organismal environmental DNA (eDNA) and microbial DNA found in the terrestrial and aquatic environments extracted from physical samples. These sequences will be used for identifying the predominant species and assessing the variability of organisms along the transect and between the soil, sediment and water environments.

Planned to be stored in <https://www.insdc.org/>

Physical object - "Material Sample "

This project will produce a set of material samples of sediment, soil, and filters from water samples. All samples will be shipped to the U.S. under an APHIS permit held by Scott Tighe.

Samples will include water - river, water - ocean; sediment - river; sediment - ocean, soil - by river.

Sample IDs: unique IDs assigned for each sampling site, recorded on metadata spreadsheet and on the sampling tubes

ISD2-Site#-Sample#

Scientific papers - "Scientific Papers"

The project's primary goal and first paper is methodological: (i) to demonstrate the value of genomic standards for both genomic research and secondary interdisciplinary research, (ii) to highlight the importance of data curation workflows that follow FAIR+CARE principles maximizing the potential for appropriate reuse of data and material samples, (iii) to illustrate the challenges that remain in implementing those standards - including ethical, legal, and social aspects, and (iv) to demonstrate some of the tools, infrastructures, and best-practices available for overcoming these challenges. This paper will include all the elements of a data paper (primary publication of the initial data collected), but with broader discussion around it as a use case for metadata standards and best practices in multi-omic sampling.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Genetic sequences from soil, sediment and concentr ...	Dataset	Unspecified	Open	Genomic Observatories MetaDatabase		Creative Commons Attribution 4.0 International	Minimum Information about any (x) Sequence (MIxS)	No	No
Material Sample	Physical object	Unspecified	Open	None specified		Creative Commons Attribution 4.0 International	None specified	No	No
Scientific Papers	Scientific papers	Unspecified	Open	None specified		None specified	None specified	No	No

Related Works

Supplemental informations

- <https://n2t.net/ark:/21547/EBW2>
- https://docs.google.com/document/d/1EYgEEUroMBiTnc5Px11_44G-B17JtTgyssyNkK3hV_U/edit#heading=h.ckwz8v2yry1q