# Host genome and microbiome sequencing data for Porites cryptic lineages in classic and extreme reefs in Palau in November 2021

**Website**: https://www.bco-dmo.org/dataset/996942
**Data Type**: Other Field Results
**Version**: 1
**Version Date**: 2026-04-17

**Project**
» Collaborative Research: How do selection, plasticity, and dispersal interact to determine coral success in warmer and more variable environments? (Palau coral selection plasticity dispersal)

| Contributors | Affiliation | Role |
|---|---|---|
| Grupstra, Carsten G.B. | Boston University (BU) | Scientist |
| Rauch, Shannon | Woods Hole Oceanographic Institution (WHOI BCO-DMO) | BCO-DMO Data Manager |

**Abstract**
This dataset contains host genomic and associated microbiome sequence data generated to investigate patterns of cryptic lineage structure, symbiotic diversity, and microbial community composition in reef-building corals (Porites spp.) across environmental gradients. Samples were collected from three classic (typical temperatures) and three extreme (higher temperatures and light attenuation) reef sites in Palau, and processed using high-throughput sequencing approaches to characterize (1) host population genomic variation and (2) taxonomic diversity of coral-associated microbial communities (Symbiodiniaceae and bacteria). Host genomic data were generated using a 2bRAD approach and analyzed to resolve lineage differentiation. Microbiome communities were characterized through amplicon sequencing, enabling assessment of microbial assemblages associated with distinct host lineages and environmental conditions.

## Table of Contents



## Coverage

**Location**: Rock Islands of Palau
**Spatial Extent**: **N**:7.36722 **E**:134.4766 **S**:7.16162 **W**:134.34697
**Temporal Extent**: 2021-11 - 2021-11

### Methods & Sampling

Colonies resembling the gross morphology of *Porites lobata* Dana, 1846 were tagged at six sites, at the Rock Islands of Palau, in November 2021 in a transect along the shoreline (N=15 per site, 90 colonies total). All colonies were sampled using a hammer and chisel between 1 and 6 meters (m) depth, with the majority between 3 and 4 m. All selected colonies were at least 1-5 m apart to reduce the risk of sampling clone mates while maximizing the probability that the colonies were exposed to similar conditions within a site. Targeted colonies were also relatively small in size (30-50 centimeters (cm)) to facilitate transportation to aquarium facilities for further analyses and experiments. The total area over which corals were collected was 250-500

square meters ($m^2$) per site. Tissue samples were taken from the center of each colony, immediately fixed in ethanol, and stored at -20 degrees Celsius (°C) (2 × 2 cm samples).

Tissue samples from all coral colonies were crushed with a sterile razor blade, and DNeasy Blood and Tissue kits (Qiagen) were used to isolate DNA from the resulting homogenate according to the manufacturer's instructions, with one modification: the lysis step was conducted overnight. Isolated DNA was then cleaned with a Zymo Clean and Concentrator kit (Zymo Research, CA). DNA was quantified using a Qubit fluorometer (Invitrogen), standardized to 25 nanograms per microliter (ng $\mu L^{-1}$), prepared for 2b-RAD sequencing according to (Wang et al., 2012), and sequenced across one lane of Illumina HiSeq 2500 using single-end 50 bp sequencing at the Tufts University Core Facility (TUCF) Genomics. Five technical replicates were included in the library preparation to aid the downstream identification of clonemates.

Photobiont communities were characterized in samples through sequencing of the internal transcribed spacer region 2 (ITS2) region using *SYM_VAR_5.8S2* and *SYM_VAR_REV* primers (Hume et al., 2015, 2018). The PCR profile included 26 cycles of 95 °C for 40 seconds, 59 °C for 2 minutes, 72 °C for 1 minute and a final extension of 72 °C for 7 minutes. A negative control was included in the initial amplification but failed to amplify, so it was not included in downstream library preparations. Successful amplifications were cleaned using the GeneJET PCR Purification kit (ThermoFisher Scientific) and a second PCR was conducted to attach Illumina MiSeq dual barcodes to the PCR product before samples were pooled. Volumes for pooling were based on the visualization of barcoded sample band intensity on a 1% agarose gel. This pool was cleaned using the GeneJET PCR Purification kit, gel extracted, and submitted for sequencing as described below.

To characterize bacterial communities, the V4 region of the 16S rRNA gene was amplified from the same samples via PCR using Hyb515f (Parada et al., 2016) and Hyb806R (Apprill et al., 2015) primers and the following PCR profile: 35 cycles of 95 °C for 40 seconds, 65 °C for 2 minutes, 72 °C for 1 minute and a final extension of 7 minutes. Subsequent PCR amplification, cleaning, dual-barcoding, and gel extraction followed the same protocol described for ITS2 with the inclusion of three negative controls, which were also submitted for sequencing. ITS2 and 16S pools were quantified and combined in a 1:3 ratio, respectively. Libraries were sequenced together on Illumina MiSeq (paired-end 250 bp) at Tufts University Core Facility (TUCF) Genomics.

**Data Processing Description**

**2bRAD Sequencing:**
Raw 2bRAD reads were deduplicated and trimmed using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit). Reads under 25 base pairs (bp) in length or with quality scores <15 were discarded. Following Rippe et al. (2021), photobiont reads were removed by discarding reads that mapped to concatenated Symbiodiniaceae genomes (*Symbiodinium* (Aranda et al., 2016)*, Breviolum* (Shoguchi et al., 2013)*, Cladocopium* (Dougan, 2020), and *Durusdinium* with Bowtie2 v2.4.2 (Langmead & Salzberg, 2012). The remaining host reads were then mapped to the *Porites lobata* genome (Noel et al., 2023). Genotyping was performed using ANGSD v0.923 (Korneliussen et al., 2014). Filters that were used across all analyses included loci that were present in ≥ 80% of individuals, and a minimum read depth of 6 across all samples. Triallelic sites were removed. Reads had a minimum quality of 25, minimum mapping quality of 20 with a strand bias p-value of 1e-5 and a heterozygosity bias p-value of 1e-5. Clones were detected using hierarchical clustering based on pairwise identity by state (IBS) with an additional minor allele frequency (MAF) filter of 0.05. Technical replicates provided the clone detection threshold, and one member of each clone pair was removed for downstream analyses.

A total of 75 samples remained after quality control and technical replicate removal. These libraries were selected for further population genomic analyses due to a higher proportion of the genome covered compared to reduced RAD samples. For all population genomic analyses an additional MAF filter of 0.05 was added, with the exception of site frequency spectrum (SFS) based analyses. Admixture was estimated using NGSadmix; admixture plots were then created using a custom R script (https://github.com/z0on/2bRAD_denovo/blob/master/admixturePlotting_v5.R). Principal Component Analysis (PCA) was conducted using a covariance matrix based on single-read resampling calculated in ANGSD. Admixture results were visualized using the K with the least cross validation error reported from ADMIXTURE. These analyses demonstrated the presence of three distinct lineages amongst our six sampling sites. $F_{ST}$ was estimated between pairs of lineages using ANGSD before and after outlier loci were removed using Bayescan (https://doi.org/10.1534/genetics.108.092221, Foll & Gaggiotti, 2008).

**Microbial community sequencing:**
Photobiont communities were characterized through sequencing of the internal transcribed spacer region 2

(ITS2) region using *SYM_VAR_5.8S2* and *SYM_VAR_REV* primers (Hume et al., 2015, 2018). The PCR profile included 26 cycles of 95 °C for 40 seconds, 59 °C for 2 minutes, 72 °C for 1 minute and a final extension of 72 °C for 7 minutes. A negative control was included in the initial amplification but failed to amplify, so it was not included in downstream library preparations. Successful amplifications were cleaned using the GeneJET PCR Purification kit (ThermoFisher Scientific) and a second PCR was conducted to attach Illumina MiSeq dual barcodes to the PCR product before samples were pooled. Volumes for pooling were based on the visualization of barcoded sample band intensity on a 1% agarose gel. This pool was cleaned using the GeneJET PCR Purification kit, gel extracted, and submitted for sequencing as described below.

To characterize bacterial communities, the V4 region of the 16S rRNA gene was amplified from the same samples via PCR using Hyb515f (Parada et al., 2016) and Hyb806R (Apprill et al., 2015) primers and the following PCR profile: 35 cycles of 95 °C for 40 seconds, 65 °C for 2 minutes, 72 °C for 1 minute and a final extension of 7 minutes. Subsequent PCR amplification, cleaning, dual-barcoding, and gel extraction followed the same protocol described for ITS2 with the inclusion of three negative controls, which were also submitted for sequencing. ITS2 and 16S pools were quantified and combined in a 1:3 ratio, respectively. Libraries were sequenced together on Illumina MiSeq (paired-end 250 bp) at Tufts University Core Facility (TUCF) Genomics.

Sequences with adaptor contamination were removed and raw 16S and ITS-2 sequences were separated based on primer sequences using bbduk following Bove et al. (2023). Raw ITS-2 reads were processed by Symportal (Hume et al., 2019) to produce defining intragenomic sequence variant (DIV) profiles for each coral colony. Two samples with <1,000 reads were removed, as well as one outlier sample with >1 million reads; the remaining samples (n = 73) had an average of ~5,500 reads per sample (min: 1,116; max: 16,931). All samples were dominated by one of ten *Cladocopium* C15 types, and four samples hosted low abundances of *Symbiodinium* A3 sequences.

**BCO-DMO Processing Description**

- Imported original file "BCO_DMO_sequencing data_final.csv" into the BCO-DMO system.
- Converted Date field to YYYY-MM format.
- Renamed fields to comply with BCO-DMO naming conventions.
- Saved the final file as "996942_v1_cryptic_porites_lineage.csv".

- Imported the supplemental file "Supplementary Datafile 3_its2-seq.csv" into the BCO-DMO system.
- Converted Date field to YYYY-MM format.
- Saved the final file as "996942_v1_supplemental_div_data.csv".

[ table of contents | back to top ]

## Related Publications

Apprill, A., McNally, S., Parsons, R., & Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquatic Microbial Ecology, 75(2), 129–137. doi:10.3354/ame01753
*Methods*

Boston University. Microbial communities associated with Porites massive lineages in classic and extreme reefs in Palau. 2024/08. In: BioProject [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; 2011-. Available from: http://www.ncbi.nlm.nih.gov/bioproject/PRJNA1154296. NCBI:BioProject: PRJNA1154296.
*IsRelatedTo*

Bove, C. B., Greene, K., Sugierski, S., Kriefall, N. G., Huzar, A. K., Hughes, A. M., Sharp, K., Fogarty, N. D., & Davies, S. W. (2023). Exposure to global change and microplastics elicits an immune response in an endangered coral. Frontiers in Marine Science, 9. https://doi.org/10.3389/fmars.2022.1037130
*Results*

Foll, M., & Gaggiotti, O. (2008). A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics, 180(2), 977–993. https://doi.org/10.1534/genetics.108.092221
*Software*

Grupstra, C. G. B., Meyer-Kaiser, K. S., Bennett, M., Andres, M. O., Juszkiewicz, D. J., Fifer, J. E., Da-Anoy, J. P.,

Gomez-Campo, K., Martinez-Rugerio, I., Aichelman, H. E., Huzar, A. K., Hughes, A. M., Rivera, H. E., & Davies, S. W. (2024). Holobiont Traits Shape Climate Change Responses in Cryptic Coral Lineages. Global Change Biology, 30(11). Portico. https://doi.org/10.1111/gcb.17578
*Results*

Hume, B. C. C., D'Angelo, C., Smith, E. G., Stevens, J. R., Burt, J., & Wiedenmann, J. (2015). Symbiodinium thermophilum sp. nov., a thermotolerant symbiotic alga prevalent in corals of the world's hottest sea, the Persian/Arabian Gulf. Scientific Reports, 5(1). https://doi.org/10.1038/srep08562
*Methods*

Hume, B. C. C., Smith, E. G., Ziegler, M., Warrington, H. J. M., Burt, J. A., LaJeunesse, T. C., Wiedenmann, J., & Voolstra, C. R. (2019). SymPortal: A novel analytical framework and platform for coral algal symbiont next-generation sequencing ITS2 profiling. Molecular Ecology Resources, 19(4), 1063–1080. Portico. https://doi.org/10.1111/1755-0998.13004
*Methods*

Hume, B. C. C., Ziegler, M., Poulain, J., Pochon, X., Romac, S., Boissin, E., de Vargas, C., Planes, S., Wincker, P., & Voolstra, C. R. (2018). An improved primer set and amplification protocol with increased specificity and sensitivity targeting the Symbiodinium ITS2 region. PeerJ, 6, e4816. Portico. https://doi.org/10.7717/peerj.4816
*Methods*

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics, 15(1). https://doi.org/10.1186/s12859-014-0356-4
*Software*

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), 357–359. doi:10.1038/nmeth.1923
*Software*

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental Microbiology, 18(5), 1403–1414. doi:10.1111/1462-2920.13023
*Methods*

Rippe, J. P., Dixon, G., Fuller, Z. L., Liao, Y., & Matz, M. (2021). Environmental specialization and cryptic genetic divergence in two massive coral species from the Florida Keys Reef Tract. Molecular Ecology, 30(14), 3468–3484. Portico. https://doi.org/10.1111/mec.15931
*Methods*

Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M., Hamada, M., Seidi, A., Fujie, M., Usami, T., Goto, H., Yamasaki, S., Arakaki, N., Suzuki, Y., Sugano, S., … Satoh, N. (2013). Draft Assembly of the Symbiodinium minutum Nuclear Genome Reveals Dinoflagellate Gene Structure. Current Biology, 23(15), 1399–1408. https://doi.org/10.1016/j.cub.2013.05.062
*Methods*

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. Nature Methods, 9(8), 808–810. doi:10.1038/nmeth.2023
*Methods*

[ table of contents | back to top ]

---

## Parameters

| Parameter | Description | Units |
|---|---|---|
| Library_Name | Name of the sample library | unitless |
| Latitude | Latitude of the sample collection site | decimal degree |
| Longitude | Longitude of the sample collection site | decimal degree |
| Date | Year and month of collection | unitless |
| Study_Accession | NCBI accession number of the sudy | unitless |
| Host_genetic_dataset_accession | NCBI accession of the overall 2bRAD dataset | unitless |
| Experiment_Accession | NCBI accession of the experiment | unitless |
| Library_accession_2BRAD | NCBI accession of the 2bRAD data for each sample | unitless |
| Microbial_community_data_accession | NCBI accession of the overall microbial sequencing dataset | unitless |
| ITS2_library_biosample | NCBI accession of the ITS-2 amplicon sequencing BioSample | unitless |
| ITS2_Forward | Filename of the ITS-2 forward read | unitless |
| ITS2_Reverse | Filename of the ITS-2 reverse read | unitless |
| Library_biosample_16S | NCBI accession of the 16S amplicon sequencing BioSample | unitless |
| Forward_16S | Filename of the 16S forward read | unitless |
| Reverse_16S | Filename of the 16S reverse read | unitless |

[ table of contents | back to top ]

## Instruments

| | |
|---|---|
| **Dataset-specific Instrument Name** | Illumina MiSeq i100 |
| **Generic Instrument Name** | Automated DNA Sequencer |
| **Dataset-specific Description** | Libraries were sequenced on an Illumina MiSeq. |
| **Generic Instrument Description** | A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences. |

| | |
|---|---|
| **Dataset-specific Instrument Name** | Illumina HiSeq 2500 |
| **Generic Instrument Name** | Automated DNA Sequencer |
| **Dataset-specific Description** | DNA was sequenced across one lane of Illumina HiSeq 2500 using single-end 50 bp sequencing. |
| **Generic Instrument Description** | A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences. |

| | |
|---|---|
| **Dataset-specific Instrument Name** | Hammer and chisel |
| **Generic Instrument Name** | Manual Biota Sampler |
| **Dataset-specific Description** | Colonies were sampled using a hammer and chisel. |
| **Generic Instrument Description** | "Manual Biota Sampler" indicates that a sample was collected in situ by a person, possibly using a hand-held collection device such as a jar, a net, or their hands. This term could also refer to a simple tool like a hammer, saw, or other hand-held tool. |

| | |
|---|---|
| **Dataset-specific Instrument Name** | Qubit 4, Invitrogen |
| **Generic Instrument Name** | Qubit fluorometer |
| **Dataset-specific Description** | DNA was quantified using a Qubit fluorometer. |
| **Generic Instrument Description** | Benchtop fluorometer. The Invitrogen Qubit Fluorometer accurately and quickly measures the concentration of DNA, RNA, or protein in a single sample. It can also be used to assess RNA integrity and quality.  Manufactured by Invitrogen, Carlsbad, CA, USA (Invitrogen is one of several brands under the Thermo Fisher Scientific corporation.) |

| | |
|---|---|
| **Dataset-specific Instrument Name** | Bibby Scientific PCRmax Alpha Cycler 4 |
| **Generic Instrument Name** | Thermal Cycler |
| **Dataset-specific Description** | Bibby Scientific tetrad of 96-well gradient Mastercyclers (PCRmax Alpha4) |
| **Generic Instrument Description** | A thermal cycler or "thermocycler" is a general term for a type of laboratory apparatus, commonly used for performing polymerase chain reaction (PCR), that is capable of repeatedly altering and maintaining specific temperatures for defined periods of time. The device has a thermal block with holes where tubes with the PCR reaction mixtures can be inserted. The cycler then raises and lowers the temperature of the block in discrete, pre-programmed steps. They can also be used to facilitate other temperature-sensitive reactions, including restriction enzyme digestion or rapid diagnostics. (adapted from http://serc.carleton.edu/microbelife/research_methods/genomics/pcr.html) |

[ table of contents | back to top ]

## Project Information

**Collaborative Research: How do selection, plasticity, and dispersal interact to determine coral success in warmer and more variable environments? (Palau coral selection plasticity dispersal)**

**Coverage**: Palauan coral reefs

**NSF Award Abstract:**
Coral reefs host thousands of marine species, help protect coastlines from storm damage, generate tourism, and house fish used for human consumption. However, corals are vulnerable to increasing water temperatures, which can lead to coral death. One way for reefs to survive in warming oceans is for corals that are well-suited to warmer waters to repopulate reefs that have less temperature-tolerant individuals. For this strategy to succeed, however, the more temperature-tolerant corals need to be able to disperse to and survive in these different environments. This project takes advantage of reef systems in the Pacific nation of Palau that naturally experience a wide range in temperatures across short geographic distances. Using cutting-edge ecological and genomic techniques, the team of investigators is directly testing whether young corals from Palau's warmest reefs can successfully be carried by ocean currents to Palau's currently cooler reefs and subsequently survive and thrive in these habitats. Given the relevance of this research for the local ecology, the team is disseminating results to the Palauan government through a written report in conjunction with Palauan scientists who are interning with the team, and to the Palauan people through public presentations. As part of this work, the investigators are maintaining a blog and are organizing a music-lecture series combining dance, music, and science to promote awareness of the coral reef crisis across English and Spanish-speaking communities in the US. Results from this project are informing restoration and conservation practices of the Coral Conservation Consortium as well as other efforts worldwide.

A major question in evolutionary biology is how plasticity and adaptation interact to influence survival under novel environments. Understanding these processes is increasingly important as rising temperatures associated with climate change influence species globally. For marine organisms with pelagic larval phases, including reef-building corals, the post-settlement period constitutes a critical bottleneck for adaptation and plasticity, with the added complexity that the conditions experienced and time spent as larvae can incur carryover effects. This project leverages reefs in Palau that span a steep environmental gradient to study how environmental variation drives selection and plasticity and to examine if dispersal between reefs limits success across habitats due to carryover effects. The investigators are testing the overarching hypothesis that corals from warmer and more variable environments are adapted to warmer temperatures and exhibit increased plasticity, but that dispersal between reefs incurs a fitness cost. The team integrates field and molecular techniques to: 1) investigate the degree of selection occurring on warmer and more variable reefs, 2) test

whether corals transplanted to more variable environments improve their thermal tolerance through developmental plasticity, and 3) examine whether delays in metamorphosis required for dispersal across reefs comes at a fitness cost due to carryover effects.

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-2048589 |
| NSF Division of Ocean Sciences (NSF OCE) | OCE-2048678 |


[ table of contents | back to top ]