# Raw amplicon, metagenomic and metatranscriptomic data from R/V JOIDES Resolution IODP-385 Guaymas Basin Tectonics and Biosphere expedition between September and November 2019

**Website**: https://www.bco-dmo.org/dataset/994734
**Data Type**: experimental
**Version**: 1
**Version Date**: 2026-03-16

**Project**
» Collaborative Research: Hydrothermal Fungi in the Guaymas Basin Hydrocarbon Ecosystem (HOTFUN)
» Collaborative Research: IODP-enabled Insights into Fungi and Their Metabolic Interactions with Other Microorganisms in Deep Subsurface Hydrothermal Sediments (IODP insights Fungi)


| Contributors | Affiliation | Role |
|---|---|---|
| Teske, Andreas | University of North Carolina at Chapel Hill (UNC-Chapel Hill) | Principal Investigator |
| Edgcomb, Virginia P. | Woods Hole Oceanographic Institution (WHOI) | Co-Principal Investigator |
| Mara, Paraskevi | Woods Hole Oceanographic Institution (WHOI) | Scientist, Contact |
| Soenen, Karen | Woods Hole Oceanographic Institution (WHOI BCO-DMO) | BCO-DMO Data Manager |


**Abstract**
International Ocean Discovery Program Expedition 385 (IODP385) drilled organic-rich sediments and intruded sills in the off-axis region and axial graben of the northern spreading segment of Guaymas Basin, Gulf of California. Guaymas Basin is characterized by high heat flow and magmatism in the form of sill intrusions into sediments, which extends tens of kilometers off axis. Sill intrusions provide transient heat sources that mobilize buried sedimentary carbon and drive hydrothermal circulation. The resulting thermal and geochemical gradients shape abundance, composition, and activity of the deep subsurface biosphere of the basin. Among the aims of IODP385 was to examine the subsurface biosphere of Guaymas Basin and its responses and adaptations to hydrothermal conditions. Using high-throughput 16S ribosomal nucleic acid sequencing we examined linkages and feedbacks between mineral composition, temperature, geochemistry, and microbial populations. The metagenomic analyses of the hydrothermally heated, sediment layers of Guaymas Basin examined the distribution and activity patterns of bacteria and archaea along thermal, geochemical and cell count gradients. Analyses of gene expression of subsurface bacteria and archaea provided insights into their physiological adaptations to in situ subsurface conditions. All available datasets can be found in the the National Center for Biotechnology Information (NCBI) GenBank database under the Bioproject accession number PRJNA909197.

## Table of Contents


- Coverage
- Dataset Description
  - Methods & Sampling
  - Data Processing Description
  - BCO-DMO Curation Notes
- Data Files
- Related Publications
- Related Datasets
- Parameters
- Instruments
- Deployments
- Project Information
- Funding


## Coverage

**Location**: Guaymas Basin, Mexico

**Spatial Extent**: **N**:27.637367 **E**:-111.13 **S**:27.12 **W**:-111.879687
**Temporal Extent**: 2019-10-01 - 2025-02-01

## Dataset Description

NSF grant OCE-2046799 and NSF grant OCE-1829903 supported salary time, lab supplies and sequencing costs

### Methods & Sampling

**Sample collection**
Sediment cores were collected during IODP Expedition 385 using the drilling vessel JOIDES Resolution. Holes at each site were first advanced using advanced piston coring (APC), then half-length APC, and then extended core barrel (XCB) coring as necessary. Temperature measurements used the advanced piston corer temperature (APCT-3) and Sediment Temperature 2 (SET2) tools8. Downhole logging conducted after coring used the triple combination and Formation MicroScanner sonic logging tool strings. After bringing core sections onto the core receiving platform of the D/V JOIDES Resolution, whole round samples for microbiology were retrieved within ~30 minutes using ethanol-cleaned spatulas. Samples for biogeochemical measurements were obtained and processed shipboard (Teske et al., 2021). Whole round samples for DNA-based studies were capped with ethanol-sterilized endcaps, transferred to the microbiology laboratory, and stored briefly at 4 °C in heat-sealed tri-foil gas-tight laminated bags flushed with nitrogen until processing. Masks, gloves and laboratory coats were worn during sample handling in the laboratory where core samples were transferred from their gas-tight bags onto sterilized foil on the bench surface inside a Table KOACH T 500-F system, which creates an ISO Class I clean air environment (Koken Ltd., Japan). In addition, the bench surface was targeted with a fanless ionizer (Winstat BF2MA, Shishido Electrostatic Co., Ltd., Japan). Within this clean space, the exterior 2 cm of the extruded core section were removed using a sterilized ceramic knife. The core interior was transferred to sterile 50-mL Falcon tubes, labeled, and immediately frozen at –80 °C for post cruise analyses. For RNA-based studies, sampling occurred immediately after core retrieval on the core receiving platform by sub-coring with a sterile, cutoff 50cc syringe into the center of each freshly cut core section targeted. These sub-cores were immediately frozen in liquid nitrogen and stored at –80 °C.

**DNA extraction and sequencing**
DNA was extracted from selected core samples using a FastDNA SPIN Kit for Soil (MP Biomedicals). Up to 5 grams of sediment were processed following a modified manufacturer's protocol (Ramirezet al., 2018). Briefly, each sediment sample was homogenized twice (vs. once that the manufacturer suggests) in Lysing Matrix E tubes for 40 seconds at speed 5.5 m/s, using the MP biomedicals bench top homogenizer equipped with 2 ml tube adaptors. Between the two homogenization rounds the samples were placed on ice for 2 minutes. After the second homogenization the samples were centrifuged at 14,000 x g for 5 minutes. For each sample, the supernatant and the top layer of the pellet was transferred to a clean 2 ml tube where proteins were precipitated by the addition of the protein precipitation solution (PPS) provided in the extraction kit. The rest of the extraction protocol followed the manufacturer's recommendations. When parallel extractions were performed, the extracts were pooled and concentrated using EMD 3kDa Amicon Ultra-0.5 ml Centrifugal Filters (Millipore Sigma). A control extraction, in which no sediment was added, was included to account for any laboratory contaminants. All libraries for metagenome sequencing (n = 29; 26 samples and 3 controls) were prepared from genomic DNA extracts that were submitted at the University of Delaware DNA Sequencing & Genotyping Center. Thirteen libraries were sequenced with NovaSeq S4 PE150 (Illumina) at the University of California, Davis Genome Center, and thirteen libraries were sequenced with NextSeq550 (Illumina) at the University of Delaware DNA Sequencing & Genotyping Center. Metagenome sequence reads were deposited to the National Center for Biotechnology Information Sequence Read Archive under access numbers SRR23614663-23614677 and SRR22580794-SRR22580807 (Bioproject PRJNA909197).

**Prokaryotic (bacterial and archaeal) 16S rRNA gene amplification and Illumina MiSeq sequencing**
The 16S rRNA gene V4/V5 hypervariable regions were targeted using the general prokaryotic primer pair 515F-Y (5′-GTGYCAGCMGCCGCGGTAA-3′; Parada et al., 2016) and 926 R (5′-CCGYCAATTYMTTTRAGTTT-3′; Quince et al., 2011) to recover 16S rRNA gene fragments of both Bacteria and Archaea. Libraries for samples 1547B-1H2, 1547B-3H2, 1545B-1H2, and 1545B-6H2 were prepared from DNA extracts by the Georgia Genomics and Bioinformatics Core (GGBC) at the University of Georgia. Libraries for all other samples and control fluid filters were prepared internally through the amplification steps described below, before sent to GGBC (for Illumina MiSeq) or the University of Delaware DNA Sequencing & Genotyping Center for final library preparation and sequencing. Illumina MiSeq overhang adapter sequences were added to locus-specific primers for use in

first round Polymerase Chain Reaction (PCR) amplifications. 16S rRNA PCR amplifications were performed for each sample (1:10 dilution) in triplicate using SpeedStar™ HS DNA Polymerase (TaKaRa) and 10X Fast Buffer I as described by the manufacturer. Thermocycling conditions were: 95 °C for 5 min; x30 (95 °C for 30 s, 60 °C for 30 s, 72 °C for 60 s); 72 °C for 5 min and a 4 °C hold. PCR amplification replicates were combined and purified with AMPure® XP beads (Beckman Coulter). Extraction kit control amplifications were attempted at both Laboratories where extractions were physically performed with only one yielding an amplicon sufficient for library construction. All libraries produced with 515F-Y/926 R amplified fragments were sequenced on an Illumina MiSeq platform with PE300 read lengths.

**Archaeal 16S rRNA gene amplification and PacBio sequencing**

To better capture the diversity of the subsurface archaeal community, we used the archaeal 16S rRNA primer set Arch25F and Arch806R that targets the V2-V4 hypervariable regions and generate a ~800 base pairs 16S rRNA amplicon. The use of larger 16S rRNA gene amplicons provides an improved basis for phylogenetic identifications, and alternate primers reduce the dependency on the extremely widespread Prokaryotic Miseq primer set (Parada et al., 2016) both factors are essential for the detection of rare or novel phylogenetic lineages41. The primer set is Arch25F (5′TCYGKTTGATCCYGSCRG 3′; Urbach et al., 2001) and Arch806R (5′GGACTACVSGGGTATCTAAT 3′; Takai et al., 2000). The 806R primer site is unusually conserved among archaea, including uncultured subsurface lineages; Teske et al., 2018). PCR reactions were performed using the SpeedSTARTM HS DNA Polymerase (TaKaRa) kit with the following modifications: each 25 µM PCR reaction contained up to 1 ng of template DNA, 2X Fast Buffer I, 2.5 mM dNTP mixture, 5 units of SpeedSTAR HS DNA Polymerase, 10 mM of each primer and DEPC water (Fisher BioReagents™) up to 25 µM. The PCR reactions were performed in an Eppendorf Mastercycler Pro S Vapoprotect (Model 6321) thermocycler with the following conditions: 95 °C for 5 min, followed by 30 cycles of 94 °C (30 s), 55 °C (30 s), 72 °C (45 s). The total volume of PCR reactions was run in 2% agarose gel (Low-EEO/Multi-Purpose/Molecular Biology Grade Fisher BioReagents™) and the correct size PCR products (~800 bp), were isolated and recovered from the gel using the Zymoclean Gel DNA Recovery Kit as instructed by the manufacturer. Libraries for PacBio sequencing were prepared from the recovered and gel purified DNA extracts at the University of Delaware DNA Sequencing & Genotyping Center.

**RNA extraction and sequencing**

Total RNA was extracted successfully from 19 sediment samples from sites U1545B-U1552B. Before each RNA extraction, all samples including a blank sample (control), were washed twice with absolute ethanol (200 proof; purity ≥ 99.5%; Thermo Scientific Chemicals), and one time with DEPC water (Fisher BioReagents) to remove hydrocarbons and other inhibitory elements present in Guyamas sediments, that without these washes, resulted in low or zero RNA yield. In brief, 13–15g of frozen sediments were transferred into UV-sterilized 50 ml Falcon tubes (RNAase/DNase free) using clean, autoclaved and ethanol-washed metallic spatulas. Each tube received an equal volume of absolute ethanol and was shaken manually for 2 min followed by 30 s of vortexing at full speed to create a slurry. Samples were transferred into an Eppendorf centrifuge (5810 R) and were centrifuged at room temperature for 2 min at 2000 rpm. The supernatant was decanted, and the ethanol wash was repeated. After decanting the supernatant of the second ethanol wash, an equal volume of DEPC water was added into each sample. Samples were manually shaken and vortexed as before to create slurry, and were transferred into the Eppendorf centrifuge (5810 R) where they were centrifuged at room temperature for 2 min at 2000 rpm. The supernatant was decanted, and each sediment sample was immediately divided into three bead-containing 15 mL Falcon tubes, provided by the PowerSoil Total RNA Isolation Kit (Qiagen). RNA was extracted as suggested by the manufacturer with the modification that the RNA extracted from the three aliquots was pooled into one RNA collection column and eluted at 30 µl final volume. All RNA extractions were performed in a UV-sterilized clean hood (two UV cycles of 15 min each) that was installed with HEPA filters. Surfaces inside the hood and pipettes were thoroughly cleaned with RNase AWAY (Thermo Scientific) before every RNA extraction and in between extraction steps. Trace DNA contaminants were removed from RNA extracts using TURBO Dnase (Thermo Fisher Scientific) and the manufacturer's protocol. Removal of DNA from the RNA extracts was confirmed with PCR reactions using the bacterial primers BACT1369F/PROK1541R (F: 5′CGGTGAATACGTTCYCGG 3′, R: 5′AAGGAGGTGATCCRGCCGCA 3′; Parada et al., 2016), targeting the small ribosomal subunit (SSU) of 16S rRNA gene. Each 25 µl PCR reaction was prepared using GoTaq G2 Flexi DNA Polymerase (Promega) and contained 0.5 U µl−1 GoTaq G2 Flexi DNA Polymerase, 1X Colorless GoTaq Flexi Buffer, 2.5 mM $MgCl_2$, (Promega) 0.4 mM dNTP Mix (Promega), 4 µM of each primer (final concentrations), and DEPC water. These PCR amplifications were performed in an Eppendorf Mastercycler Pro S Vapoprotect (Model 6321) thermocycler with following conditions: 94 °C for 5 min, followed by 35 cycles of 94 °C (30 s), 55 °C (30 s), and 72 °C (45 s). The PCR reaction products were run in 2% agarose gels (Low-EEO/Multi-Purpose/Molecular Biology Grade Fisher BioReagents) to confirm absence of DNA products. RNA quantification (ng µl−1) was performed using Qubit RNA High Sensitivity (HS), Broad Range (BR), and Extended Range (XR) Assay Kits, (Invitrogen).

Amplified cDNAs from the DNA-free RNA extracts were prepared using the Ovation RNA-Seq System V2

(Tecan) following manufacturer's suggestions. cDNAs were submitted to the Georgia Genomics and Bioinformatics Core for library preparation and sequencing using NextSeq 500 PE 150 High Output (Illumina). The sequencing of the cDNA library from the control sample was unsuccessful as it failed to generate any sequences that met the length criterion of 300-400 base pairs.

**Data Processing Description**

**16S rRNA marker gene analyses**
Sequenced reads for 16S rRNA gene fragments were analyzed with the QIIME2 pipeline for paired-end (Illumina MiSeq) and single-end (PacBio) sequencing employing the DADA2 denoising method for amplicon sequence variant (ASV) construction(Boylen et al., 2019; Callahan et al., 2016). ASVs in our samples that were matches to ASVs in either kit or drilling fluid controls were removed within the QIIME2 pipeline. In addition, prior to downstream analyses, we manually remove any remaining ASVs in our data sets that were taxonomically annotated to known contaminants of human and terrestrial origin and kit contaminants (Salter et al., 2014). Drilling fluid taxa removed from the final data set include: Mycobacteriales and Propionibacteriales (Actinobacteria), Staphylococcales (Firmicutes), Paceibacterales, UBA1400, UBA9983_A, and Microgenomatia (Patescibacteria), and Burkholderiales and Pseudomonadales (Gammaproteobacteria). Taxonomy was assigned using a trained classifier (q2-feature-classifier (Bokulich et al., 2018) with the SILVA_132_QIIME_release or SILVA_v138.1_release(Quast et al., 2013; Yilmaz et al., 2014) as a reference database for Illumina MiSeq or PacBio sequences respectively.

Illumina MiSeq 16S rRNA gene reads are deposited into the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under access numbers SRR23641000-SRR23641053. The PacBio reads are deposited under SRA access numbers SRR23604162-SRR23604206 (Bioproject PRJNA909197).

**Metatranscriptome data analyses**
Raw sequencing reads were trimmed to remove adapters and low-quality bases using fastp (v0.23.2) (Chen et al., 2018) with parameters (-q 20 -u 20 -l 50 -w 16 -5 -M 30 -g -D --detect_adapter_for_pe --dup_calc_accuracy 6). We used Trinity (v2.14.0) (Grabherr et al., 2011) to assemble the 19 metatranscriptomes with default settings. Trinity generated 640,136 assembled metatranscripts with size > 165 bp. We performed DIAMOND (v.2.0.7) BLASTx (Buchfink et al., 2021) against NCBI-NR database (release date: 2022-12-04) to provide functional and taxonomic annotations on the assembled metatranscriptomes. Because the control sample failed to generate sequences that met the minimum length criterion, the annotated transcripts with e-values > 1e–5 were manually curated to remove possible contaminants by creating an in-house database that contained putative contaminant species (taxa identified as potential kit contaminants and human pathogens; e.g., Salter et al., 2014). Transcripts with > 90% similarity to the in-house database over >50% of contig length were removed for downstream analyses. This process removed 8,301 transcripts (8,301/640,136; ~1.2%). The remaining decontaminated assembled transcripts were processed with Prodigal (v2.6.3; Hyatt et al., 2010) to predict gene and protein sequences. CD-hit Fu et al., 2012) (v. 4.8.1; -c 0.95 -aS 0.9 -n 10) was used to cluster genes and to remove redundancy. For functional annotation, KofamScan (v.1.3.0) (Aramaki et al., 2019 and GhostKOALA (v.2.2) Kanehisa et al., 2016) were used to assign orthologs (KOs) to protein sequences using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. DIAMOND (v.2.0.7) BLASTp (Buchfink et al., 2021) (v2.0.15.153, -e 1e–5 --more-sensitive) was used to search against NCBI-NR database (release date: 2022-12-04).

To gain a more detailed insight on the function and taxonomy of the expressed genes, the decontaminated and non-redundant assembled transcripts were re-run with DIAMOND BLASTx (Buchfink et al., 2021) (v2.0.15.153, -e 1e–5 --more-sensitive) against NCBI-NR database (release date: 2022-12-04). The BLASTx results with e-values > 1e–5 were manually curated for expression of genes involved in methane, nitrogen, sulfur and folate metabolisms, carbon fixation, C1 from folate, glycine cleavage system, DNA maintenance and repair, RNA modifications, proteostasis, proteolysis, arsenic detoxification, cadmium and copper transport, tungsten-containing aldehyde ferredoxin oxidoreductases, circadian rhythm, ferredoxins, rubredoxins, signal recognition particle protein, archaeal flagellin, von Willebrand type A domains, sporulation, Ni-Fe hydrogenases, haloacid reductive dehalogenases. We recognize that any automated and manual pipeline that is used to assign gene function has the caveat that publicly available databases may contain some protein sequences that have not been functionally validated on the bench.

The expression level of each transcript was estimated in units of transcripts per million (TPM) using Salmon (v1.9.0, --meta) (Patro et al., 2017). The TPM values of all transcripts annotated to same gene were summed

and were added to a value of 1 (to avoid zeros) and normalized using log2-transformation.

Metatranscriptome reads were deposited to the National Center for Biotechnology Information Sequence Read Archive under accession numbers SRR22580929-SRR22580947 (BioProject PRJNA909197).

### BCO-DMO Curation Notes

* Merged 2 submitted datasets into 1 file.
* Treated empty strings, "not applicable" and "nd" as missing values
* Split lat_lon field into latitude and longitude
* Set longitude values to negative for west and positive for east
* Reformatted date fields Collection_Date, ReleaseDate, and create_date from MM-DD-YY to YYYY-MM-DD format
* Renamed fields with spaces or special characters to underscore-separated names
* Deleted duplicate fields Accession_number_SRA and Location
* Corrected malformed value "0.607.64" to "0.60764" in Bytes field
* Removed " Gb" suffix from Bytes field values
* Added constant field BioProject with value "PRJNA909197"
* Reordered all 39 fields into specified column order
* Removed processing notes that were not dataset specific

All items were discussed and approved by the submitter

[ table of contents | back to top ]

## Data Files

| File | |
|---|---|
| **994734_v1_metagenome.csv** | (Comma Separated Values (.csv), 87.43 KB) MD5:bc6c26aac51324eccd97e901296dccbd |
| Primary data file for dataset ID 994734, version 1 | |

[ table of contents | back to top ]

## Related Publications

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. Nature Methods, 11(11), 1144–1146. https://doi.org/10.1038/nmeth.3103
*Methods*

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., & Ogata, H. (2019). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics, 36(7), 2251–2252. https://doi.org/10.1093/bioinformatics/btz859
*Methods*

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., Huttley, G. A., & Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome, 6(1). https://doi.org/10.1186/s40168-018-0470-z
*Methods*

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. doi:10.1093/bioinformatics/btu170
*Methods*

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
*Software*

Buchfink, B., Reuter, K., & Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. Nature Methods, 18(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x
*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
*Software*

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics, 36(6), 1925–1927. https://doi.org/10.1093/bioinformatics/btz848
*Software*

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560
*Software*

Chklovski, A., Parks, D. H., Woodcroft, B. J., & Tyson, G. W. (2023). CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nature Methods, 20(8), 1203–1212. https://doi.org/10.1038/s41592-023-01940-w
*Software*

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics, 28(23), 3150–3152. doi:10.1093/bioinformatics/bts565
*Methods*

Geller-McGrath, D., Konwar, K. M., Edgcomb, V. P., Pachiadaki, M., Roddy, J. W., Wheeler, T. J., & McDermott, J. E. (2024). Predicting metabolic modules in incomplete bacterial genomes with MetaPathPredict. ELife, 13. CLOCKSS. https://doi.org/10.7554/elife.85749 https://doi.org/10.7554/eLife.85749
*Methods*

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29(7), 644–652. doi:10.1038/nbt.1883
*Software*

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics, 11(1). doi:10.1186/1471-2105-11-119
*Methods*

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. Journal of Molecular Biology, 428(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006
*Software*

Kang, D. D., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ, 3, e1165. doi:10.7717/peerj.1165
*Software*

Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., … Lam, T.-W. (2016). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods, 102, 3–11. doi:10.1016/j.ymeth.2016.02.020
*Software*

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2015). 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*

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods, 14(4), 417–419. doi:10.1038/nmeth.4197
*Methods*

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research, 41(D1), D590–D596. doi:10.1093/nar/gks1219
*Methods*

Quince, C., Lanzen, A., Davenport, R. J., & Turnbaugh, P. J. (2011). Removing Noise From Pyrosequenced Amplicons. BMC Bioinformatics, 12(1). https://doi.org/10.1186/1471-2105-12-38
*Methods*

Ramírez, G. A., Graham, D., & D'Hondt, S. (2018). Influence of commercial DNA extraction kit choice on prokaryotic community metrics in marine sediment. Limnology and Oceanography: Methods, 16(9), 525–536. Portico. https://doi.org/10.1002/lom3.10264
*Methods*

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., … Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology, 12(1). doi:10.1186/s12915-014-0087-z
*Methods*

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153
*Software*

Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., & Banfield, J. F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nature Microbiology, 3(7), 836–843. https://doi.org/10.1038/s41564-018-0171-1
*Methods*

Takai, K., & Horikoshi, K. (2000). Rapid Detection and Quantification of Members of the Archaeal Community by Quantitative PCR Using Fluorogenic Probes. Applied and Environmental Microbiology, 66(11), 5066–5072. https://doi.org/10.1128/aem.66.11.5066-5072.2000 https://doi.org/10.1128/AEM.66.11.5066-5072.2000
*Methods*

Teske, A., & Sørensen, K. B. (2007). Uncultured archaea in deep marine subsurface sediments: have we caught them all? The ISME Journal, 2(1), 3–18. https://doi.org/10.1038/ismej.2007.90
*Methods*

Teske, A., Lizarralde, D., Höfig, T. W., Aiello, I. W., Ash, J. L., Bojanova, D. P., Buatier, M. D., Edgcomb, V. P., Galerne, C. Y., Gontharet, S., Heuer, V. B., Jiang, S., Kars, M. A. C., Khogenkumar Singh, S., Kim, J., Koornneef, L. M. T., Marsaglia, K. M., Meyer, N. R., Morono, Y., … Zhuang, G. (2021). Expedition 385 summary. Guaymas Basin Tectonics and Biosphere. Internet Archive. https://doi.org/10.14379/iodp.proc.385.101.2021
*Methods*

Vasimuddin, Md., Misra, S., Li, H., & Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), 314–324. https://doi.org/10.1109/ipdps.2019.00041 https://doi.org/10.1109/IPDPS.2019.00041
*Methods*

Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2015). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics, 32(4), 605–607. https://doi.org/10.1093/bioinformatics/btv638
*Methods*

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., … Glöckner, F. O. (2013). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Research, 42(D1), D643–D648. doi:10.1093/nar/gkt1209
*Methods*

Zhou, Z., Tran, P. Q., Breister, A. M., Liu, Y., Kieft, K., Cowley, E. S., Karaoz, U., & Anantharaman, K. (2022). METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. Microbiome, 10(1). https://doi.org/10.1186/s40168-021-01213-8
*Methods*

[ table of contents | back to top ]

## Related Datasets

### IsRelatedTo

Woods Hole Oceanographic Institution. (2022). IODP expedition 385 Guaymas Basin Tectonics and Biosphere metatranscriptome raw data [Data set]. National Center for Biotechnology Information, BioProject.

## Parameters

| Parameter | Description | Units |
|---|---|---|
| BioProject | NCBI BioProject accession ID | unitless |
| SRA_Run_ID | NCBI SRA sample accession ID | unitless |
| BioSample | NCBI Biosample accession ID | unitless |
| AssayType | Type of sequencing performed (e.g., metagenome, amplicon) | unitless |
| Assay | Type of sequencing performed | unitless |
| AvgSpotLen | Calculated average read length (in base pairs) across sequencing spots | unitless |
| Bases | Total number of sequenced bases | unitless |
| BioSampleModel | The BioSample package/model selected | unitless |
| Bytes | Size of the sequencing data for the run | unitless |
| CenterName | Name of sequencing or submitting center | unitless |
| CollectionDate | Collection date of sample | unitless |
| SRA_Experiment_ID | SRA experiment accession | unitless |
| geo_loc_name_country | Country where the sample was collected | unitless |
| geo_loc_name_country_continent | Continent corresponding to the sample location | unitless |
| geo_loc_name | Geographic location of the origin of the sample | unitless |
| Instrument | Sequencing platform used | unitless |

| | | |
|---|---|---|
| Latitude | Latitude of sampling location, south is negative | decimal degrees |
| Longitude | Longitude of sampling location, west is negative | decimal degrees |
| LibraryName | Submitter-provided name identifying the sequencing library | unitless |
| LibraryLayout | single or paired end sequencing reads | unitless |
| LibrarySelection | Selection used for sequencing library | unitless |
| LibrarySource | Biological source of the library material | unitless |
| Organism | Organism name by submitter | unitless |
| Platform | Sequencing platform manufacturer | unitless |
| ReleaseDate | Date the data became publicly available | unitless |
| CreateDate | Date the record was created in NCBI systems | unitless |
| SampleName | Name identifying the biological sample | unitless |
| IsolationSource | Environment from which the sample originated | unitless |
| SampCollectDevice | Device used to collect the sample | unitless |
| Depth_mbsf | depth of sediment sample below seafloor | meter (m) |
| AmplificationPrimers | Primer sequences used during PCR amplification | unitless |
| elev | Elevation or depth relative to sea level | meter (m) |
| env_broad_scale | Broad-scale environmental context | unitless |
| env_local_scale | Local-scale environmental context | unitless |
| env_medium | Material displaced by the entity at time of sampling | unitless |

| | | |
|---|---|---|
| rel_to_oxygen | Oxygen relationship of the environment | unitless |

[ table of contents | back to top ]

## Instruments

| | |
|---|---|
| **Dataset-specific Instrument Name** | Illumina NovaSeq S4 PE150 |
| **Generic Instrument Name** | Automated DNA Sequencer |
| **Dataset-specific Description** | High-throughput sequencing platform from Illumina |
| **Generic Instrument Description** | A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences. |

| | |
|---|---|
| **Dataset-specific Instrument Name** | Illumina NextSeq550 PE150x2 |
| **Generic Instrument Name** | Automated DNA Sequencer |
| **Dataset-specific Description** | NextSeq550 PE150x2 (Illumina): It is part of the NextSeq 550 System, which is a high-throughput sequencing platform from Illumina that can perform paired-end sequencing with read lengths up to 150 base pairs (bp) (max read length of 2 x 150 bp). |
| **Generic Instrument Description** | A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences. |

| | |
|---|---|
| **Dataset-specific Instrument Name** | Illumina MiSeq platform |
| **Generic Instrument Name** | Automated DNA Sequencer |
| **Dataset-specific Description** | MiSeq PE300: refers to a sequencing technique run on an Illumina MiSeq platform where the machine generates paired-end reads, with each read being 300 base pairs long, meaning it sequences both ends of a DNA fragment, producing two 300 base pair reads for each fragment. Paired-end (PE): This indicates that the machine will sequence both the forward and reverse strands of a DNA fragment, providing more information about the sequence compared to single-end reads. 300 bp: Each individual read will be 300 base pairs long. |
| **Generic Instrument Description** | A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences. |

[ table of contents | back to top ]

## Deployments

**IODP-385**

| | |
|---|---|
| **Website** | https://www.bco-dmo.org/deployment/869491 |
| **Platform** | R/V JOIDES Resolution |
| **Report** | https://doi.org/10.14379/iodp.pr.385.2020 |
| **Start Date** | 2019-09-15 |
| **End Date** | 2019-11-15 |
| **Description** | Guaymas Basin Tectonics and Biosphere - International Ocean Discovery Program Expedition 385, General information: https://iodp.tamu.edu/scienceops/expeditions/guaymas_basin_tectonics_bio... |

[ table of contents | back to top ]

## Project Information

**Collaborative Research: Hydrothermal Fungi in the Guaymas Basin Hydrocarbon Ecosystem (HOTFUN)**

**Coverage**: Guaymas Basin, Gulf of CA, Mexico

NSF Award Abstract:
Fungi that can derive energy from chemicals, yet consume other organisms or organic material to obtain carbon have been reported from diverse marine subsurface samples, including from hundreds of meters below the seafloor. Evidence exists that Fungi are active in subsurface marine sediments globally, yet there is a dearth of knowledge on their role in the marine subsurface, and specifically on their role(s) in hydrocarbon degradation within deep-sea sediments. This team is isolating a broad collection of environmentally relevant filamentous Fungi and yeasts from hydrothermally-influenced and hydrocarbon-rich seep sediments of Guaymas Basin using high-throughput culture-based approaches. They aim to reveal the diversity of Fungi and Bacteria in these hydrothermal sediments, how temperature and hydrocarbon composition shape their distribution, and how Fungi cooperate to enhance the degradation of hydrocarbons by Bacteria. By hosting six undergraduates through the WHOI Summer Student Fellows program and the Woods Hole Partnership Education Program, the project contributes to increasing diversity in marine science by offering opportunities for promising undergraduates from disadvantaged populations. High school students are involved in summer projects and in intensive summer workshops. One postdoc, a graduate student, and two Research Associates are supported, and international collaborations are strengthened. The postdoc and graduate student are gaining valuable cruise-based experience. An e-lecture on Fungi and their role(s) in biodegradation of hydrocarbons will be made publicly available by the end of the project. Fungal isolates with accompanying information will be secured in a reference culture collection for long-term storage and are available to any interested researcher throughout the project.

The PIs are isolating a broad collection of environmentally relevant filamentous Fungi and yeasts from hydrothermally-influenced and hydrocarbon-rich seep sediments of Guaymas Basin using high-throughput culture-based approaches, with the aim to reveal their ability to degrade individual hydrocarbons under in situ pressures and temperatures. Culture independent methods marker gene analyses are used to characterize in situ fungal and bacterial diversity and to examine how temperature and hydrocarbon composition shape fungal community composition and distribution. Traditional and comprehensive two-dimensional gas chromatographic analyses are used to examine the complexities and subtle changes in inventories of hydrocarbons within sediment cores, and provide evidence for in situ microbial alteration of individual hydrocarbons. Incubation experiments are used to test the ability of fungal isolates to utilize different hydrocarbons as a sole or auxiliary carbon source under in situ pressures and temperatures and their ability to stimulate biodegradation of hydrocarbons by hydrocarbon-degrading bacteria. Expressed genes within these incubation studies tell us how Fungi and Bacteria couple metabolisms to increase overall specificity and extent of biodegradation of hydrocarbons.

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.

**Collaborative Research: IODP-enabled Insights into Fungi and Their Metabolic Interactions with Other Microorganisms in Deep Subsurface Hydrothermal Sediments (IODP insights Fungi)**

**Website**: https://www2.whoi.edu/site/edgcomblab/research/deep-marine-subsurface-eukaryotes/

**Coverage**: Guaymas Basin, Gulf of California, Mexico

## NSF Award Abstract

The marine subsurface is one of the least explored habitats on Earth. International Ocean Discovery Program (IODP) Expedition 385 drilled into the seafloor in Guaymas Basin, Mexico, and was the first to drill directly into subsurface sediments and sediment-hosted basalt sill intrusions of an active hydrothermal basin. This expedition provides a direct microbiological window into a deep hydrothermal biosphere across an active plate spreading center where complex hydrocarbons are generated by heating of buried organic matter under high temperature and pressure. Mounting evidence suggests that Fungi constitute an active and ecologically important fraction of the subsurface biosphere community. This is especially true in organic-rich continental margin sediments that are ideal for colonization by aerobic and anaerobic Fungi, where fungal activities may contribute significantly to nutrient cycling. Major knowledge gaps in our knowledge of subsurface microbiota preclude our ability to estimate their full impact: how active Fungi are distributed along temperature and depth gradients, the range of substrates utilized by active cells, and how Fungi may cooperate with bacteria in degradation of complex organic matter, including hydrocarbons. Fungi are known to participate in degradation of refractory organics and cycling of metals and to produce novel metabolites with interesting properties. This project informs us on origins of different lineages of microbial life on Earth, the extent of marine subsurface carbon cycling, limits of life, how life adapts to environmental change, and the potential for Fungi to accelerate the biodegradation of complex hydrocarbons. Given the extent of the potential subsurface biosphere, Fungi likely play an important role in global nutrient cycling. The culture collection of fungal isolates created by this project will be available for exploration of their ecology and novel properties by other interested researchers, and may also yield insights into basal fungal lineages. These biogeochemical and potential evolutionary outcomes are of great interest to other research disciplines, educators, and students alike. The project's K-16 education program capitalizes on programs aimed at increasing involvement of under-represented undergraduate populations in research. High school students, undergraduates, a graduate student, and postdoc are involved in the research. An art-in-science project with a local high school is being displayed at the community library along with education materials on marine Fungi and their ecological roles.

This project examines how abundance, diversity and distribution of Fungi and co-inhabiting bacteria and archaea changes in subsurface sediment samples exhibiting a wide range of in situ temperatures and pressure, what the active fraction of cells is along these gradients, and whether/how Fungi impact carbon cycling in this biosphere by interacting metabolically with bacteria to break down hydrocarbon substrates. The project is assessing the activities of in situ microorganisms in this active hydrothermal subsurface biosphere using a cutting-edge combination of molecular approaches and culture-based studies of enrichments and microbial isolates applied to an extensive collection of samples from 8 sites in Guaymas Basin varying in temperature profile, presence of old, buried magmatic sills, and geochemical conditions. The investigators are examining 1) marker genes and metagenomes of sorted active cells using new bioorthogonal non-canonical amino acid tagging (BONCAT) approaches, 2) the distribution of bacterial, archaeal, and fungal cells and their marker genes along depth and geochemical gradients using microscopy, 'meta-omics' and lipid biomarker analyses, 3) substrate usage by fungal isolates, 4) metabolite pools, nutrients, and hydrocarbons with depth, and 5) fungal metabolism of complex organics (and syntrophies between Fungi and Bacteria) using time-course stable isotope probing of RNA from culture-based studies coupled with analyses of expressed genes and pools of metabolites.

## Broader Impacts

The proposed project can transform our understanding of microbial life in the sedimented marine subsurface biosphere because active mycobiota would have implications for deep carbon budgets. Our culture collection is estimated to generate hundreds of new strains of fungi that will be available for exploration of their ecology and novel properties by interested researchers, and may also yield insights into basal fungal lineages. These biogeochemical and potential evolutionary outcomes are of great interest to other research disciplines, educators, and students alike. Our proposed K-16 education program capitalizes on programs aimed at increasing involvement of under-represented undergraduate populations in research. High school students, undergraduates (4 per year), a graduate student, and postdoc will be involved. The PI and an art teacher at a local high school will teach an art-in-science unit. The product (a large quilt of art inspired by fungal cultures)

will be displayed at the community library along with education materials on marine fungi and their ecological roles. The project involves two international collaborators, and will partially support 4 principal investigators at three institutions, including two early career researchers, Roland Hatzenpichler and Paraskevi Mara.

**Additional Shipboard Data from the Expedition**

The IODP page for Guaymas Expedition 385 contains multiple site chapters, one for each drilling site (U1545-U1552); each site chapter contains sections about petrology, sedimentology, porewater chemistry, etc.; data tables are embedded into the site chapters. The citation and access DOI for this resource is as follows:

Teske, A., Lizarralde, D., & Höfig, T. W. (2021). Guaymas Basin Tectonics and Biosphere. Proceedings of the International Ocean Discovery Program. https://doi.org/10.14379/iodp.proc.385.2021

[ table of contents | back to top ]

## Funding


| Funding Source | Award |
|---|---|
| NSF Division of Ocean Sciences (NSF OCE) | OCE-1829903 |
| NSF Division of Ocean Sciences (NSF OCE) | OCE-2046799 |


[ table of contents | back to top ]