Standard Metadata				
GSC MIMARKS: Genomic Standards Consortium marker gene sequence survey				
SRA: Sequence Read Archive				
banzai: Banzai Pipeline				
biom: Post Pipeline Processing				
Field Name	**Example**	**Description**	**Source**	**Use**
order	1		banzai fields required for processing	
sample_name	canon15c11_7_edna		banzai	banzai, SRA
library	H1		banzai	banzai
library_tag_combo	H1_GGTCTAGGTCTA		banzai	banzai
tag_sequence	GGTCTAGGTCTA		banzai	banzai
primer_sequence_F	GGWACWGGWTGAACWGTWTAYCCYCC		banzai	banzai
primer_sequence_R	TAAACTTCAGGGTGACCAAAAAATCA		banzai	banzai
date_PCR	20160629		banzai	
sample_type	environmental, blank, positive, negative	environmental = true sample, blank = blanks, EB and CB, positive = mock or artificial communities, negative = NTC	banzai	banzai
locus	CO1		banzai	banzai
tag_number	1		banzai	
R1	canon15c11-7-edna_S1_L001_R1_001.fastq.gz	Raw sequence fastq read1 file	banzai	SRA
R2	canon15c11-7-edna_S1_L001_R2_001.fastq.gz	Raw sequence fastq read2 file	banzai	SRA
year	2015		banzai	biom
month	9		banzai	biom
day	30		banzai	biom

Standard Metadata				
GSC MIMARKS: Genomic Standards Consortium marker gene sequence survey				
SRA: Sequence Read Archive				
banzai: Banzai Pipeline				
biom: Post Pipeline Processing				
Field Name	**Example**	**Description**	**Source**	**Use**
replicate	A or B or C or 1 or 2 etc	Required for SRA submission as each sample name must be unique.	SRA	SRA
env_biome	marine biome (ENVO:00000447)	Biomes are defined based on factors such as plant structures, leaf types, plant spacing, and other factors like climate. Biome should be treated as the descriptor of the broad ecological context of a sample. Examples include: desert, taiga, deciduous woodland, or coral reef. EnvO (v 2013-06-14) terms can be found via the link: www.environmentontology.org/Browse-EnvO	*GSC MIMARKS Mandatory field	SRA
env_feature	coastal water (ENVO:00001250)	Environmental feature level includes geographic environmental features. Compared to biome, feature is a descriptor of the more local environment. Examples include: harbor, cliff, or lake. EnvO (v 2013-06-14) terms can be found via the link: www.environmentontology.org/Browse-EnvO	*GSC MIMARK	SRA
env_material	waterborne particulate matter (ENVO:01000436)	The environmental material level refers to the material that was displaced by the sample, or material in which a sample was embedded, prior to the sampling event. Environmental material terms are generally mass nouns. Examples include: air, soil, or water. EnvO (v 2013-06-14) terms can be found via the link: www.environmentontology.org/Browse-EnvO	*GSC MIMARK	SRA
env_package	mimarks survey	MIGS/MIMS/MIMARKS extension for reporting of measurements and observations obtained from one or more of the environments where the sample was obtained. All environmental packages listed here are further defined in separate subtables. By giving the name of the environmental package, a selection of fields can be made from the subtables and can be reported	*GSC MIMARK	SRA
project_name	MBON, M2W, A2W	Formerly the field "seq_for"; adopting GSC identifier. Name of the project within which the sequencing was organized	*GSC MIMARK	biom

Standard Metadata				
GSC MIMARKS: Genomic Standards Consortium marker gene sequence survey				
SRA: Sequence Read Archive				
banzai: Banzai Pipeline				
biom: Post Pipeline Processing				
Field Name	**Example**	**Description**	**Source**	**Use**
geo_loc_name	USA:California:Monterey Bay	The geographical origin of the sample as defined by the country or sea name followed by specific region name. Country or sea names should be chosen from the INSDC country list (http://insdc.org/country.html), or the GAZ ontology (v 1.512) (http://purl.bioontology.org/ontology/GAZ)	*GSC MIMARK	SRA
seq_meth	NGS Illumina Miseq	Sequencing method used; e.g. Sanger, pyrosequencing, ABI-solid	*GSC MIMARK	SRA
investigation_type	eukaryote, bacteria_archaea, plasmid, virus, organelle, metagenome,mimarks-survey or mimarks-specimen	Nucleic Acid Sequence Report is the root element of all MIGS/MIMS compliant reports as standardized by Genomic Standards Consortium. This field is either eukaryote,bacteria,virus,plasmid,organelle, metagenome,mimarks-survey, or mimarks-specimen	*GSC MIMARK	SRA
SAMPLING_campaign	MBTS or CANON	User defined	TARA	
SAMPLING_project	CANON	Adopting the TARA style of identifying collection metadata.	BOG	biom
SAMPLING_date_time	9/30/2015 19:00	Adopting the TARA style of identifying collection metadata.	BOG	biom
SAMPLING_platform	WESTERN FLYER	Adopting the TARA style of identifying collection metadata.	BOG	biom
SAMPLING_cruise	CANON15	Adopting the TARA style of identifying collection metadata.	BOG	biom
SAMPLING_station_number	11	Adopting the TARA style of identifying collection metadata.	BOG	biom
SAMPLING_station	5E		TARA	biom
SAMPLING_dec_lat	36.9962	Adopting the TARA style of identifying collection metadata.	BOG	biom
SAMPLING_dec_lon	-122.2367	Adopting the TARA style of identifying collection metadata.	BOG	biom
SAMPLING_bottle	7		TARA	biom
depth	40		*GSC MIMARK	SRA, biom
SAMPLING_real_depth_m	40.369	Adopting the TARA style of identifying collection metadata.	BOG	biom
SAMPLING_PI	Francisco_Chavez	Adopting the TARA style of identifying collection metadata.	BOG	
SAMPLING_institute	MBARI	Adopting the TARA style of identifying collection metadata.	BOG	
description	CN15	User defined	BOG	biom

Standard Metadata				
GSC MIMARKS: Genomic Standards Consortium marker gene sequence survey				
SRA: Sequence Read Archive				
banzai: Banzai Pipeline				
biom: Post Pipeline Processing				
Field Name	**Example**	**Description**	**Source**	**Use**
samp_collection_device	CTD/SBE911_Rosette_ with_Niskin_bottle	The method or device employed for collecting the sample	GSC MIMARKS	SRA
temp	13.5022	AtlantOS Essential Ocean Variable (EOV)	GSC MIMARKS	biom
salinity	33.3471	AtlantOS EOV	GSC MIMARKS	biom
chlorophyll	0.32485	AtlantOS EOV	GSC MIMARKS	biom
pressure_dbar	40.674	AtlantOS EOV	banzai	biom
nitrate	7.854	AtlantOS EOV	GSC MIMARKS	biom
diss_oxygen	4.47095	AtlantOS EOV	GSC MIMARKS	biom
samp_vol_we_dna_ext	1000ml		GSC MIMARKS	SRA
samp_filter_size_ext	0.22um		BOG	
samp_filter_ext_type	Poretics		BOG	
samp_store_temp	-80C		GSC MIMARKS	SRA
sequencing_facility	Stanford		BOG	
nucl_acid_ext	dx.doi.org/10.17504/proto cols.io.mvzc676	nucleic acid extraction. Link to a literature reference, electronic resource or a standard operating procedure (SOP)	GSC MIMARKS	
nucl_acid_amp	dx.doi.org/10.17504/pr otocols.io.mv2c68e	nucleic acid amplification. Link to a literature reference, electronic resource or a standard operating procedure (SOP)	GSC MIMARKS	
target_gene	18S	same as locus, GSC mandatory field	*GSC MIMARK	SRA
lat_lon	36.9962 N 122.2367 W	The geographical origin of the sample as defined by latitude and longitude. The values should be reported in decimal degrees and in WGS84 system	*GSC MIMARK	SRA
collection_date	2015-09-30	The time of sampling, either as an instance (single point in time) or interval. In case no exact time is available, the date/time can be right truncated i.e. all of these are valid times: 2008-01-23T19:23:10+00:00; 2008-01-23T19:23:10; 2008-01-23; 2008-01; 2008; Except: 2008-01; 2008 all are ISO8601 compliant	*GSC MIMARK	SRA

Standard Metadata				
GSC MIMARKS: Genomic Standards Consortium marker gene sequence survey				
SRA: Sequence Read Archive				
banzai: Banzai Pipeline				
biom: Post Pipeline Processing				
Field Name	**Example**	**Description**	**Source**	**Use**
pcr_primers	pcr primers	This field is a duplicate for the primer_sequence F&R. It is included as a GSC MIMARKS identifier. PCR primers that were used to amplify the sequence of the targeted gene, locus or subfragment. This field should contain all the primers used for a single PCR reaction if multiple forward or reverse primers are present in a single PCR reaction. The primer sequence should be reported in uppercase letters	GSC MIMARKS	
mid	multiplex identifiers	This field is a duplicate for the tag_sequence. It is included as a GSC MIMARKS identifier. Molecular barcodes, called Multiplex Identifiers (MIDs), that are used to specifically tag unique samples in a sequencing run. Sequence should be reported in uppercase letters	GSC MIMARKS	
adapters	adapters	Adapters provide priming sequences for both amplification and sequencing of the sample-library fragments. Both adapters should be reported; in uppercase letters	GSC MIMARKS	
Biom Metadata				
biom metadata includes standard metadata. Mostly setting from the pipeline parameter file.				
Field Name	**Example**	**Description**	**Source**	**Use**
submitted_to_insdc	0		GSC	
lib_const_meth	paired-end		GSC	
assembly	PEAR		GSC	
PEAR_Quality_Threshold	19		banzai	
PEAR_UNCALLEDMAX	0		banzai	
PEAR_TEST	1		banzai	
PEAR_PVALUE	0.01		banzai	
PEAR_SCORING	2		banzai	
PEAR_min_seq_length	75		banzai	
BLAST_BLAST_DB	'/MBON/blastdb/nt/nt'		banzai	
BLAST_PERCENT_IDENTITY	80		banzai	

Standard Metadata				
GSC MIMARKS: Genomic Standards Consortium marker gene sequence survey				
SRA: Sequence Read Archive				
banzai: Banzai Pipeline				
biom: Post Pipeline Processing				
Field Name	**Example**	**Description**	**Source**	**Use**
BLAST_WORD_SIZE	20		banzai	
BLAST_EVALUE	1.00E-05		banzai	
BLAST_MAXIMUM_MATCHES	100		banzai	
BLAST_culling_limit	20		banzai	
CLUSTER_OTUS	YES		banzai	
cluster_method	swarm"		banzai	
cluster_radius	1		banzai	
min_OTU_abun	0.005		banzai	
REMOVE_HOMOPOLYMERS	NO		banzai	
HOMOPOLYMER_MAX	7		banzai	
PRIMER_MISMATCH_PROPORTION	0.1		banzai	
remove_chimeras	NO		banzai	
Perform_Expected_Error_Filter	YES"		banzai	
Max_Expected_Errors	0.5		banzai	
BLAST	blastn: 2.2.31+		banzai	
swarm	[Oct 22 2015 08:54:34]		banzai	
vsearch	, 15.5GB RAM, 12 cores		banzai	
cutadapt	1.8.3		banzai	
python	Python 2.7.5		banzai	
PEAR	PEAR v0.9.6 [January 15, 2015] - [+bzlib +zlib]		banzai	
seqtk	Version: 1.0-r82-dirty		banzai	
R	R version 3.4.3 (2017-11-30) -- "Kite-Eating Tree"		banzai	

Standard Metadata				
GSC MIMARKS: Genomic Standards Consortium marker gene sequence survey				
SRA: Sequence Read Archive				
banzai: Banzai Pipeline				
biom: Post Pipeline Processing				
Field Name	**Example**	**Description**	**Source**	**Use**
annot_source	MEGAN6	For cases where annotation was provided by a community jamboree or model organism database rather than by a specific submitter	GSC	

A