Supporting Information for: An Expanded Methanosphere: Methane Oxidation and Methanotrophs Surrounding Cold Seeps Across Oxygen Gradients on the Southern California Margin

Emily Klonicki-Ference[1*], Daniel R. Utter[2], Kira Homola[1], John S. Magyar[2], Victoria J. Orphan[2], David W. Caress[3], Jennifer B. Paduan[3], Lisa Levin[4], Tina Treude[1,5*]

[1]Department of Earth, Planetary, and Space Sciences, University of California Los Angeles, Los Angeles, CA, USA

[2]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA,
USA

[3] Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing CA USA

[4]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA

[5]Department of Atmospheric and Oceanic Sciences, University of California Los Angeles, Los Angeles, CA, USA

*Correspondence: Tina Treude, ttreude@g.ucla.edu; Emily Klonicki-Ference, eklonicki@g.ucla.edu

Table S1. Geographic coordinates, depths, and two-dimensional distances between sequential waypoints for two surveyed locations: DMS (CTD3) and SMM (CTD7). The last waypoint in each section does not have a distance to the next waypoint, as it is the final point in the survey sequence.

Location	Waypoint	Latitude (dec. deg.)	Longitude (dec. deg.)	Depth (m)	Distance to next Waypoint (m)
DMS	1	32.9055	-117.7827	1015	101.6
DMS	2	32.9046	-117.7825	1018	48.1
DMS	3	32.9042	-117.7823	1018	56.2
DMS	4	32.9037	-117.7822	1007	93.1
DMS	5	32.9029	-117.7819	1024	1836.7
DMS	6	32.8868	-117.7773	1045	103.7
DMS	7	32.8859	-117.7770	1045	45.3
DMS	8	32.8855	-117.7769	1045	48.1
DMS	9	32.8851	-117.7767	1045	101.6
DMS	10	32.8842	-117.7765	1045	N/A
SMM	1	33.7988	-118.6449	816	72.9
SMM	2	33.7991	-118.6456	812	38.7
SMM	3	33.7992	-118.6460	809	29.9
SMM	4	33.7993	-118.6463	803	29.9
SMM	5	33.7994	-118.6466	800	21.6
SMM	6	33.7995	-118.6468	801	29.9
SMM	7	33.7996	-118.6471	805	21.6
SMM	8	33.7997	-118.6473	806	72.9
SMM	9	33.8000	-118.6480	810	81.2
SMM	10	33.8003	-118.6488	810	N/A

Table S2. Environmental and microbiological characteristics at different depths across the three seep sites: Santa Monica Mound (SMM), Del Mar Seep (DMS), and Lasuen Knoll (LK). Variables include depth, first-order rate constant (k), methane turnover time, methane (CH_4) and oxygen (O_2) concentrations, temperature, pmoA gene copy abundance (log scale), and fluorescence as a proxy for microbial biomass.

Site	Depth (m)	k	Turnover Time (y)	CH_4 (nM)	O_2 (μM)	Temp. °C	Log copies pmoA L^{-1}	Fluorescence (μg L^{-1})
SMM	2	0.001256	2.18	82	253	19.8	3.49	0.3989
SMM	200	0.000272	10.1	68	82	9.62	5.21	0.0084
SMM	400	0.000156	17.5	74	28	8.15	N/A	0.0244
SMM	500	0.001282	2.14	77	18	7.21	N/A	0.0261
SMM	580	0.004633	0.591	111	11	6.42	5.84	0.0373
SMM	600	0.003128	0.876	122	11	6.33	N/A	0.0351
SMM	703	0.011327	0.242	172	5.9	5.63	N/A	0.0343
SMM	752	0.01777	0.154	191	5.4	5.50	6.03	0.0372
SMM	791	0.018501	0.148	235	3.8	5.32	N/A	0.0417
SMM	798	0.006782	0.404	293	3.5	5.31	5.57	0.0443
DMS	2	0.001588	1.73	47	231	20.35	4.65	0.043
DMS	290	0.000142	19.3	80	56	8.95	N/A	0.016
DMS	541	0.000491	5.58	79	12	6.72	5.560	0.0212
DMS	790	0.001025	2.67	66	10	5.19	N/A	0.0355
DMS	940	0.000428	6.41	63	14	4.49	5.61	0.0379
DMS	989	0.000509	5.38	66	15	4.35	N/A	0.026
DMS	1008	0.002133	1.28	75	20	4.13	5.47	0.0402
DMS	1017	0.002624	1.04	73	22	3.97	5.07	0.0329
DMS	1026	0.001065	2.57	131	22	3.96	5.07	0.0389
DMS	1031	0.000869	3.15	113	23	3.96	N/A	0.0341
LK	3	0.003812	0.719	33	232	20.4	N/A	0.1178
LK	51	0.000525	5.22	47	233	12.8	N/A	0.7355
LK	101	0.001743	1.57	54	145	10.6	N/A	0.1266
LK	201	0.004869	0.563	62	95	9.30	5.12	0.0116
LK	291	0.000188	14.6	76	56	8.94	N/A	0.0227
LK	341	0.001174	2.33	66	45	8.47	5.32	0.0255
LK	360	0.000909	3.01	214	39	8.46	N/A	0.0217
LK	370	0.001412	1.94	89	35	8.46	5.20	0.0202
LK	378	0.001152	2.38	135	36	8.43	N/A	0.0204

Table S3. Log-transformed qPCR-derived concentrations of the particulate methane monooxygenase gene (pmoA) in water samples collected at the Del Mar Seep, Santa Monica Mound, and Lasuen Knoll. A given sample is labeled by depth, CTD, Waypoint (WP), and or *Alvin* (A) cast.

Site	CTD or Alvin Dive	Sample	Log pmoA copies/L	Standard Deviation
Del Mar	AD5193	AD5193-A - 1022m	4.53	0.16
Del Mar	AD5194	AD5194-A - 1022m	4.70	0.08
Del Mar	AD5194	AD5194-B - 1019m	2.50	0.29
Del Mar	2	CTD2 - 800m	5.82	0.01
Del Mar	2	CTD2 - 1034m	5.57	0.03
Del Mar	3	CTD3 WP1 - 1015m	5.25	0.14
Del Mar	3	CTD3 WP3 - 1017m	5.46	0.02
Del Mar	3	CTD3 WP5 - 1024m	5.45	0.03
Del Mar	3	CTD3 WP8 - 1045m	5.29	0.06
Del Mar	4	CTD4 - 2m	4.65	0.01
Del Mar	4	CTD4 - 541m	5.60	0.08
Del Mar	4	CTD4 - 940m	5.61	0.19
Del Mar	4	CTD4 - 1008m	5.47	0.03
Del Mar	4	CTD4 - 1026m	5.07	0.09
Del Mar	4	CTD4 - 1031m	5.07	0.10
Santa Monica	AD5197	AD5197-A - 802m	5.23	0.13
Santa Monica	AD5197	AD5197-B - 800m	4.42	0.14
Santa Monica	AD5198	AD5198-A - 816m	4.77	0.19
Santa Monica	AD5198	AD5198-B - 801m	5.30	0.29
Santa Monica	AD5199	AD5199-A - 816m	5.00	0.21
Santa Monica	6	CTD6 - 819m	5.81	0.16
Santa Monica	7	CTD7 WP1 - 816m	5.88	0.03
Santa Monica	7	CTD7 WP4 - 803m	6.05	0.02
Santa Monica	7	CTD7 WP6 - 801m	5.91	0.05
Santa Monica	7	CTD7 WP8 - 806m	5.93	0.00
Santa Monica	7	CTD7 WP10 - 810m	5.91	0.09
Santa Monica	8	CTD8 - 2m	3.49	0.00
Santa Monica	8	CTD8 - 200m	5.21	0.00
Santa Monica	8	CTD8 - 580m	5.84	0.05
Santa Monica	8	CTD8 - 752m	6.03	0.05
Santa Monica	8	CTD8 - 798m	5.57	0.09
Lasuen	AD5201	AD5201-A - 310m	4.99	0.06

Lasuen	11	CTD11 - 201m	5.12	0.19
Lasuen	11	CTD11 - 341m	5.32	0.01
Lasuen	11	CTD11 - 370m	5.20	0.06
Lasuen	11	CTD11 - 382m	5.15	0.02

51

16S rRNA processing via DADA2 for Klonicki et al. 2026

Marine methane (CH4) seeps are dynamic biogeochemical systems that regulate carbon and sustain high-biomass communities through microbial CH4 oxidation. While most CH4 is consumed anaerobically in sediments, a fraction escapes into the water column, where aerobic methanotrophs act as a biological filter limiting atmospheric flux. However, the spatial extent of CH4 influence beyond active seep zones remains poorly constrained, with implications for deep-sea food webs and carbon cycling. We investigated microbial CH4 turnover and methanotroph distribution across three seep sites on the Southern California margin (Del Mar (1020 m), Santa Monica Mound (800 m), and Lasuen Knoll (400 m)) focusing on extent of horizontal transport, presence of vertical gradients, and oxygen controls. Using radiotracer (3H-CH4) incubations, CH4 concentration profiles, 16S rRNA gene sequencing, and particulate methane monooxygenase (pmoA) gene quantification, we characterized CH4-fueled processes along vertical and lateral transects, including near-bottom waters sampled via HOV Alvin.

16S rRNA raw reads can be found on NCBI under BioProject PRJNA1479037. Briefly, samples were collected from Niskin bottles (from the HOV Alvin or CTD casts from the R/V *Atlantis*) onto sterivex filters. DNA was obtained with enzymatic lysis followed by phenol/chloroform extraction and the V4V5 region was amplified with 515f and 926r primers with NEB Q5 HotStart polymerase for 32 cycles. Amplicons were barcoded with Illumina adapters in a second PCR reaction and sequenced on an Illumina MiSeq instrument (2x250bp).

Following are the steps to generate amplicon sequence variants (ASVs) from the raw data using DADA2.

Cutadapt

How this works: the `ls to_upload/...` bit points to the raw data to be analyzed (here `to_upload` is the name of the samples at `PRJNA1479037`).

```
MY_FILES=$(ls to_upload/*R1*)
echo -e "Today we are analyzing:\n\n$MY_FILES"
echo ""
mkdir -p fastq_cutadapt  # to make the directory where cutadapt-processed files will go

for R1 in $MY_FILES; do
R2=$(echo $R1 | sed 's/_R1_/_R2_/')
cutadapt -a ^GTGYCAGCMGCCGCGGTAA...AAACTYAAAKRAATTGRCGG -A ^CCGYCAATTYMTTTRAGTTT...TTACCGCGGCKGCTGRCAC -j 12 -
done
```

Main DADA2 workflow

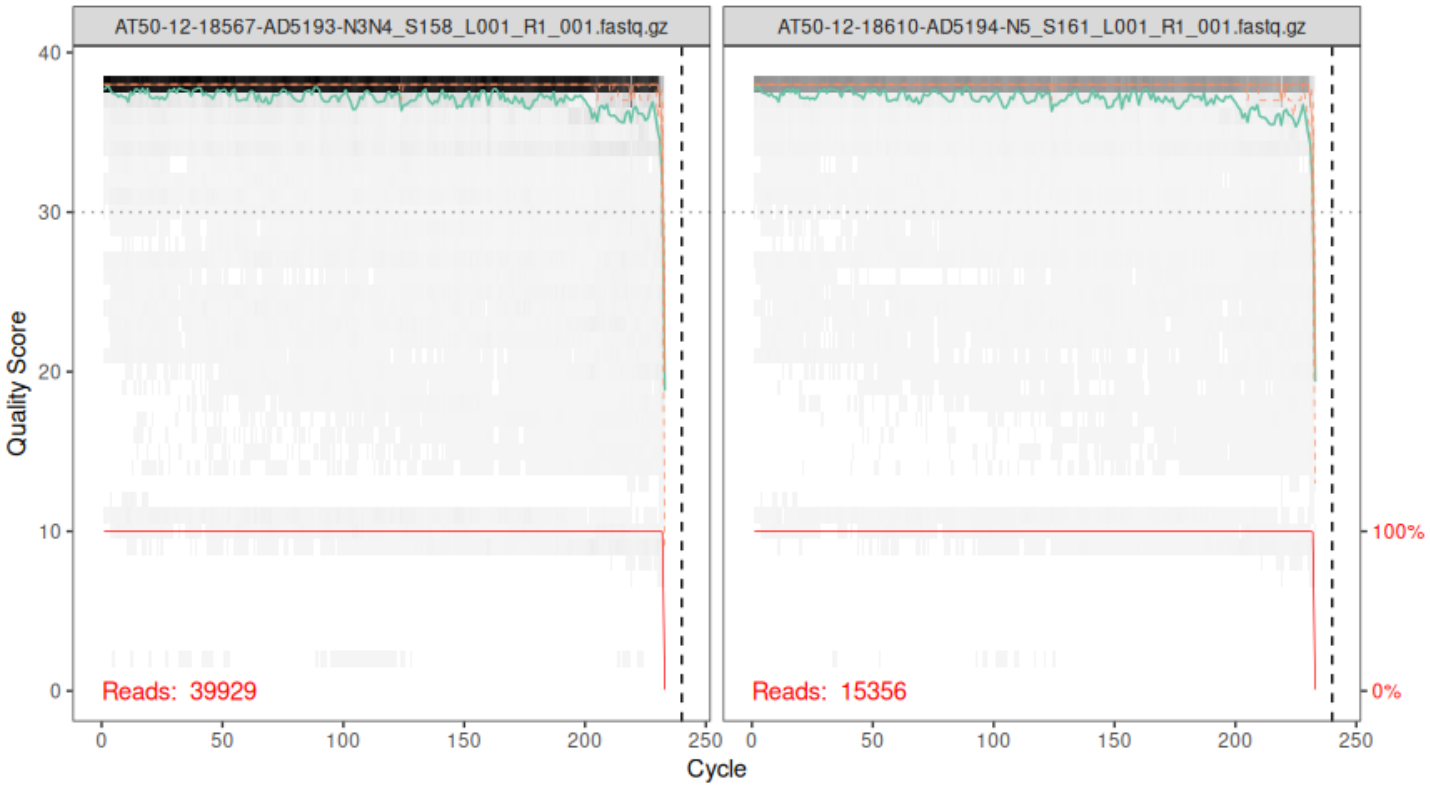
Reading in data Read in data. The last `gsub` line sets the sample names by trimming off the general filename suffixes (`S00_L001_R2.fastq.gz`).

```
library(dada2)
library(ggplot2)
library(reshape2)
library(dplyr)

forward_reads <- list.files("fastq_cutadapt", pattern="R1_001", full.names = T)
reverse_reads <- list.files("fastq_cutadapt", pattern="R2_001", full.names = T)

# this is names of files where filtered forward reads will be saved to, they're not filtered yet
filtered_forward_reads <- gsub("_001.fastq","_filtered.fastq", forward_reads)
```

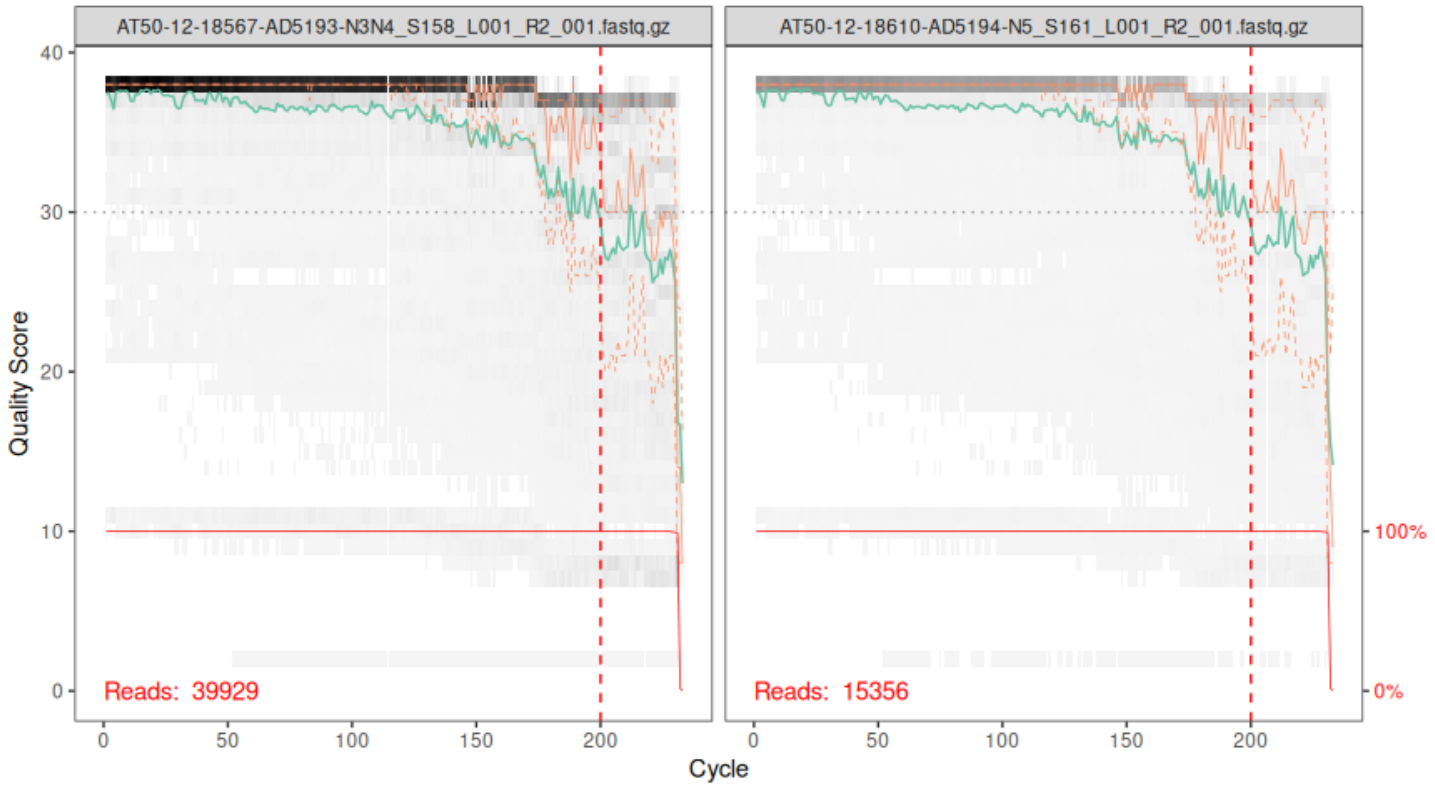
```
filtered_reverse_reads <- gsub("_001.fastq","_filtered.fastq", reverse_reads)

# set up sample names
samples <- gsub("^.*/","",
                gsub("_S[0-9]*_L.*$", "", filtered_forward_reads))
```

Can plot a few to get an idea of what quality scores look like:

```
plotQualityProfile(forward_reads[1:min(2, length(forward_reads))]) +
  geom_vline(xintercept = 240, col='black', lty=2) +
  geom_hline(yintercept = 30, col='grey60', lty=3)
```




```
plotQualityProfile(reverse_reads[1:min(2, length(forward_reads))]) +
  geom_vline(xintercept = 200, col='red', lty=2) +
  geom_hline(yintercept = 30, col='grey60', lty=3)
```



Filter & error rates Here we used 230bp forward / 200bp reverse as this gave sufficient overlap while meeting approximately Q30 for 230 on R1 which was as far as R1 could go with 2x250 reads.

```
filtered_out <- filterAndTrim(forward_reads, filtered_forward_reads,
                reverse_reads, filtered_reverse_reads, maxEE=c(2,2),
                rm.phix=TRUE, minLen=150, truncLen=c(230,200))

#cbind(filtered_out, perc_kept = filtered_out[,2]/filtered_out[,1])
```

Estimate error models & denoise (make ASVs) Using the filtered reads generated in the last step, this next chunk will estimate the error rates from our data. Sequences are then dereplicated which `dada()` uses with the error model to define ASVs and discard erroneous variants.

```
err_forward_reads <- learnErrors(filtered_forward_reads, multithread=8)
err_reverse_reads <- learnErrors(filtered_reverse_reads, multithread=8)

derep_forward <- derepFastq(filtered_forward_reads, verbose=FALSE)
derep_reverse <- derepFastq(filtered_reverse_reads, verbose=FALSE)

# the sample names in these objects are initially the file names, this sets them to the sample names
names(derep_forward) <- gsub("^.*/","", gsub("_S[0-9]*_.*$", "", filtered_forward_reads))
names(derep_reverse) <- gsub("^.*/","",gsub("_S[0-9]*_.*$", "", filtered_forward_reads))

# actual DADA2 denoising step
dada_forward <- dada(derep_forward, err=err_forward_reads, pool="pseudo", multithread=8)
dada_reverse <- dada(derep_reverse, err=err_reverse_reads, pool="pseudo", multithread=8)
```

Merge and chimera cleaning This step merges the denoised forward and reverse reads.

```
merged_amplicons <- mergePairs(dada_forward, derep_forward, dada_reverse, derep_reverse,
                               trimOverhang=TRUE, minOverlap=12)
seqtab <- makeSequenceTable(merged_amplicons)
```

```
#print("Merged fraction:")
#rowSums(seqtab) / rowSums(makeSequenceTable(dada_forward))

seqtab.nochim <- removeBimeraDenovo(seqtab, verbose = T, method = 'consensus')

#print("Non-chimeric fraction:")
#sum(seqtab.nochim)/sum(seqtab)
```

Summarize read retention by step

Tabulate and plot the fate of reads throughout DADA2 (this makes a good supplemental figure):

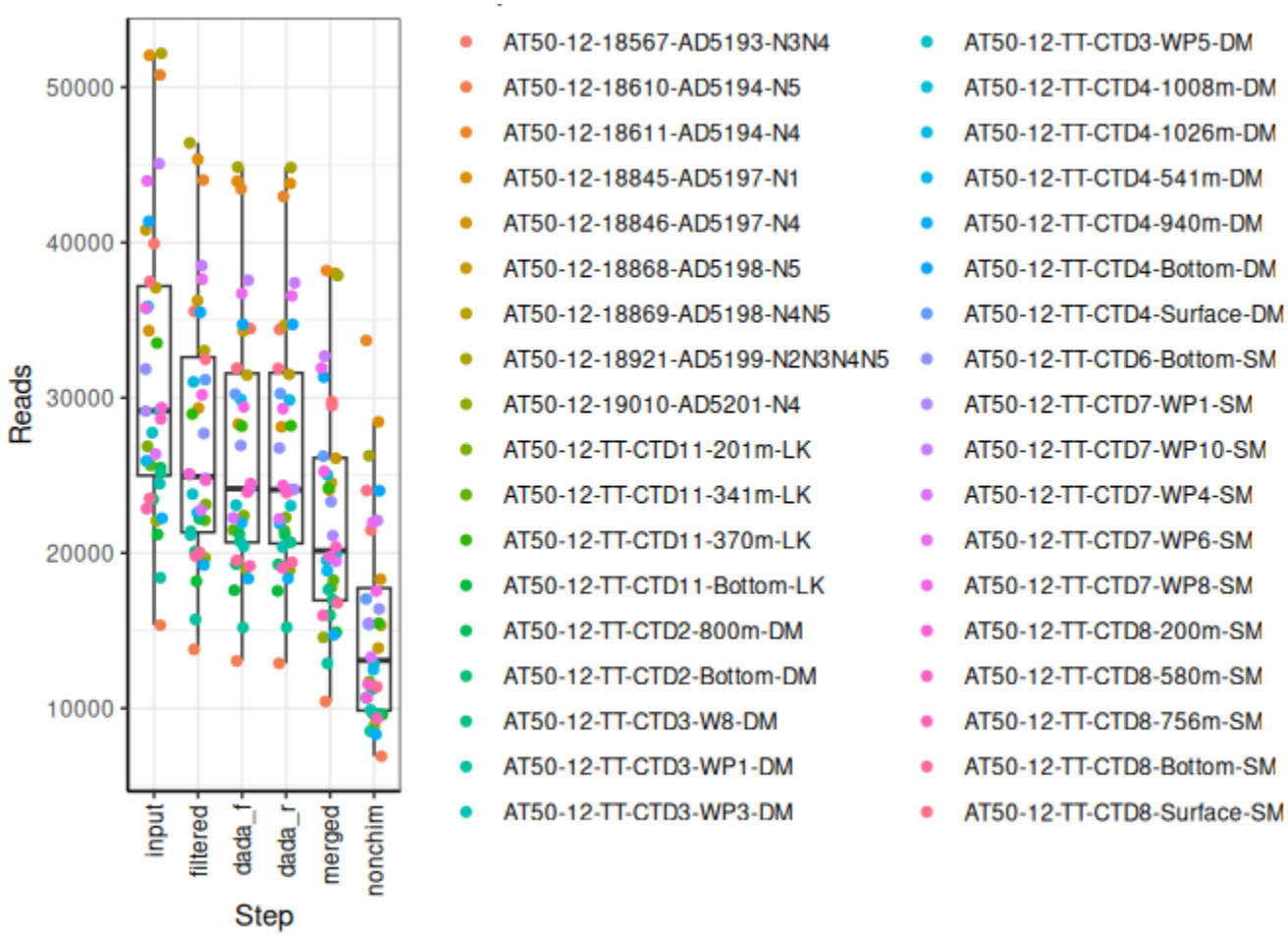
```
getN <- function(x) sum(getUniques(x))

summary_tab <- data.frame(row.names=samples, samp=samples, input=filtered_out[,1],
               filtered=filtered_out[,2], dada_f=sapply(dada_forward, getN),
               dada_r=sapply(dada_reverse, getN), merged=sapply(merged_amplicons, getN),
               nonchim=rowSums(seqtab.nochim),
               final_perc_reads_retained=round(rowSums(seqtab.nochim)/filtered_out[,1]*100, 1))

summary_tab
#write.table(summary_tab, "read-count-tracking.tsv", quote=FALSE, sep="\t", col.names=NA)

print(summary_tab)
```

```
##                                                samp input filtered
## AT50-12-18567-AD5193-N3N4         AT50-12-18567-AD5193-N3N4 39929    35569
## AT50-12-18610-AD5194-N5             AT50-12-18610-AD5194-N5 15356    13790
## AT50-12-18611-AD5194-N4             AT50-12-18611-AD5194-N4 50794    44023
## AT50-12-18845-AD5197-N1             AT50-12-18845-AD5197-N1 52057    45358
## AT50-12-18846-AD5197-N4             AT50-12-18846-AD5197-N4 34319    29338
## AT50-12-18868-AD5198-N5             AT50-12-18868-AD5198-N5 40810    36263
## AT50-12-18869-AD5198-N4N5         AT50-12-18869-AD5198-N4N5 37096    33044
## AT50-12-18921-AD5199-N2N3N4N5 AT50-12-18921-AD5199-N2N3N4N5 52182    46416
## AT50-12-19010-AD5201-N4             AT50-12-19010-AD5201-N4 22079    19685
## AT50-12-TT-CTD11-201m-LK           AT50-12-TT-CTD11-201m-LK 26873    23131
## AT50-12-TT-CTD11-341m-LK           AT50-12-TT-CTD11-341m-LK 25646    22109
## AT50-12-TT-CTD11-370m-LK           AT50-12-TT-CTD11-370m-LK 33525    28955
## AT50-12-TT-CTD11-Bottom-LK       AT50-12-TT-CTD11-Bottom-LK 21200    18179
## AT50-12-TT-CTD2-800m-DM             AT50-12-TT-CTD2-800m-DM 25504    22213
## AT50-12-TT-CTD2-Bottom-DM         AT50-12-TT-CTD2-Bottom-DM 23465    20092
## AT50-12-TT-CTD3-W8-DM                 AT50-12-TT-CTD3-W8-DM 25168    21397
## AT50-12-TT-CTD3-WP1-DM               AT50-12-TT-CTD3-WP1-DM 18412    15726
## AT50-12-TT-CTD3-WP3-DM               AT50-12-TT-CTD3-WP3-DM 24468    21154
## AT50-12-TT-CTD3-WP5-DM               AT50-12-TT-CTD3-WP5-DM 27753    23795
## AT50-12-TT-CTD4-1008m-DM           AT50-12-TT-CTD4-1008m-DM 29182    25058
## AT50-12-TT-CTD4-1026m-DM           AT50-12-TT-CTD4-1026m-DM 35887    31024
## AT50-12-TT-CTD4-541m-DM             AT50-12-TT-CTD4-541m-DM 25933    22606
## AT50-12-TT-CTD4-940m-DM             AT50-12-TT-CTD4-940m-DM 22242    19231
## AT50-12-TT-CTD4-Bottom-DM         AT50-12-TT-CTD4-Bottom-DM 41365    35509
## AT50-12-TT-CTD4-Surface-DM       AT50-12-TT-CTD4-Surface-DM 35771    31166
## AT50-12-TT-CTD6-Bottom-SM         AT50-12-TT-CTD6-Bottom-SM 31848    27704
## AT50-12-TT-CTD7-WP1-SM               AT50-12-TT-CTD7-WP1-SM 29142    24804
## AT50-12-TT-CTD7-WP10-SM             AT50-12-TT-CTD7-WP10-SM 45077    38511
## AT50-12-TT-CTD7-WP4-SM               AT50-12-TT-CTD7-WP4-SM 26370    22781
## AT50-12-TT-CTD7-WP6-SM               AT50-12-TT-CTD7-WP6-SM 43968    37639
## AT50-12-TT-CTD7-WP8-SM               AT50-12-TT-CTD7-WP8-SM 35800    30189
## AT50-12-TT-CTD8-200m-SM             AT50-12-TT-CTD8-200m-SM 29371    25091
## AT50-12-TT-CTD8-580m-SM             AT50-12-TT-CTD8-580m-SM 28645    24682
## AT50-12-TT-CTD8-756m-SM             AT50-12-TT-CTD8-756m-SM 22853    19806
```

```
## AT50-12-TT-CTD8-Bottom-SM         AT50-12-TT-CTD8-Bottom-SM 23519    20041
## AT50-12-TT-CTD8-Surface-SM       AT50-12-TT-CTD8-Surface-SM 37498    32479
##                               dada_f dada_r merged nonchim
## AT50-12-18567-AD5193-N3N4      34454  34405  29757   21481
## AT50-12-18610-AD5194-N5        13048  12895  10441    6912
## AT50-12-18611-AD5194-N4        43459  42972  38197   33688
## AT50-12-18845-AD5197-N1        43951  43802  38023   28446
## AT50-12-18846-AD5197-N4        28310  28126  24081   18310
## AT50-12-18868-AD5198-N5        34310  34675  26099   13877
## AT50-12-18869-AD5198-N4N5      31468  31509  24528   15343
## AT50-12-18921-AD5199-N2N3N4N5  44858  44832  37878   26271
## AT50-12-19010-AD5201-N4        18983  18893  14567    8903
## AT50-12-TT-CTD11-201m-LK       22407  22287  17789    9742
## AT50-12-TT-CTD11-341m-LK       21485  21440  18253   11720
## AT50-12-TT-CTD11-370m-LK       28175  28199  24175   15475
## AT50-12-TT-CTD11-Bottom-LK     17602  17559  14895    9595
## AT50-12-TT-CTD2-800m-DM        21230  21176  16987    9592
## AT50-12-TT-CTD2-Bottom-DM      19292  19271  15996    9636
## AT50-12-TT-CTD3-W8-DM          20770  20691  17645   11230
## AT50-12-TT-CTD3-WP1-DM         15201  15214  12894    8528
## AT50-12-TT-CTD3-WP3-DM         20435  20385  16844    9892
## AT50-12-TT-CTD3-WP5-DM         23085  23046  19905   12877
## AT50-12-TT-CTD4-1008m-DM       24114  24060  19527   10692
## AT50-12-TT-CTD4-1026m-DM       29887  29859  25043   15405
## AT50-12-TT-CTD4-541m-DM        21946  21908  18863   12490
## AT50-12-TT-CTD4-940m-DM        18347  18360  14730    8323
## AT50-12-TT-CTD4-Bottom-DM      34734  34727  31335   24014
## AT50-12-TT-CTD4-Surface-DM     30231  30264  26242   17042
## AT50-12-TT-CTD6-Bottom-SM      26929  26767  23313   16399
## AT50-12-TT-CTD7-WP1-SM         24186  24107  21125   15471
## AT50-12-TT-CTD7-WP10-SM        37572  37405  32696   22103
## AT50-12-TT-CTD7-WP4-SM         22270  22193  19468   13291
## AT50-12-TT-CTD7-WP6-SM         36718  36562  31915   21975
## AT50-12-TT-CTD7-WP8-SM         29415  29298  25263   17555
## AT50-12-TT-CTD8-200m-SM        24470  24349  19765   10673
## AT50-12-TT-CTD8-580m-SM        23923  23883  20390   11547
## AT50-12-TT-CTD8-756m-SM        19174  19048  15984    9349
## AT50-12-TT-CTD8-Bottom-SM      19523  19401  16793   11385
## AT50-12-TT-CTD8-Surface-SM     31912  31895  29487   24028
##                               final_perc_reads_retained
## AT50-12-18567-AD5193-N3N4                          53.8
## AT50-12-18610-AD5194-N5                            45.0
## AT50-12-18611-AD5194-N4                            66.3
## AT50-12-18845-AD5197-N1                            54.6
## AT50-12-18846-AD5197-N4                            53.4
## AT50-12-18868-AD5198-N5                            34.0
## AT50-12-18869-AD5198-N4N5                          41.4
## AT50-12-18921-AD5199-N2N3N4N5                      50.3
## AT50-12-19010-AD5201-N4                            40.3
## AT50-12-TT-CTD11-201m-LK                           36.3
## AT50-12-TT-CTD11-341m-LK                           45.7
## AT50-12-TT-CTD11-370m-LK                           46.2
## AT50-12-TT-CTD11-Bottom-LK                         45.3
## AT50-12-TT-CTD2-800m-DM                            37.6
## AT50-12-TT-CTD2-Bottom-DM                          41.1
## AT50-12-TT-CTD3-W8-DM                              44.6
## AT50-12-TT-CTD3-WP1-DM                             46.3
## AT50-12-TT-CTD3-WP3-DM                             40.4
## AT50-12-TT-CTD3-WP5-DM                             46.4
## AT50-12-TT-CTD4-1008m-DM                           36.6
```

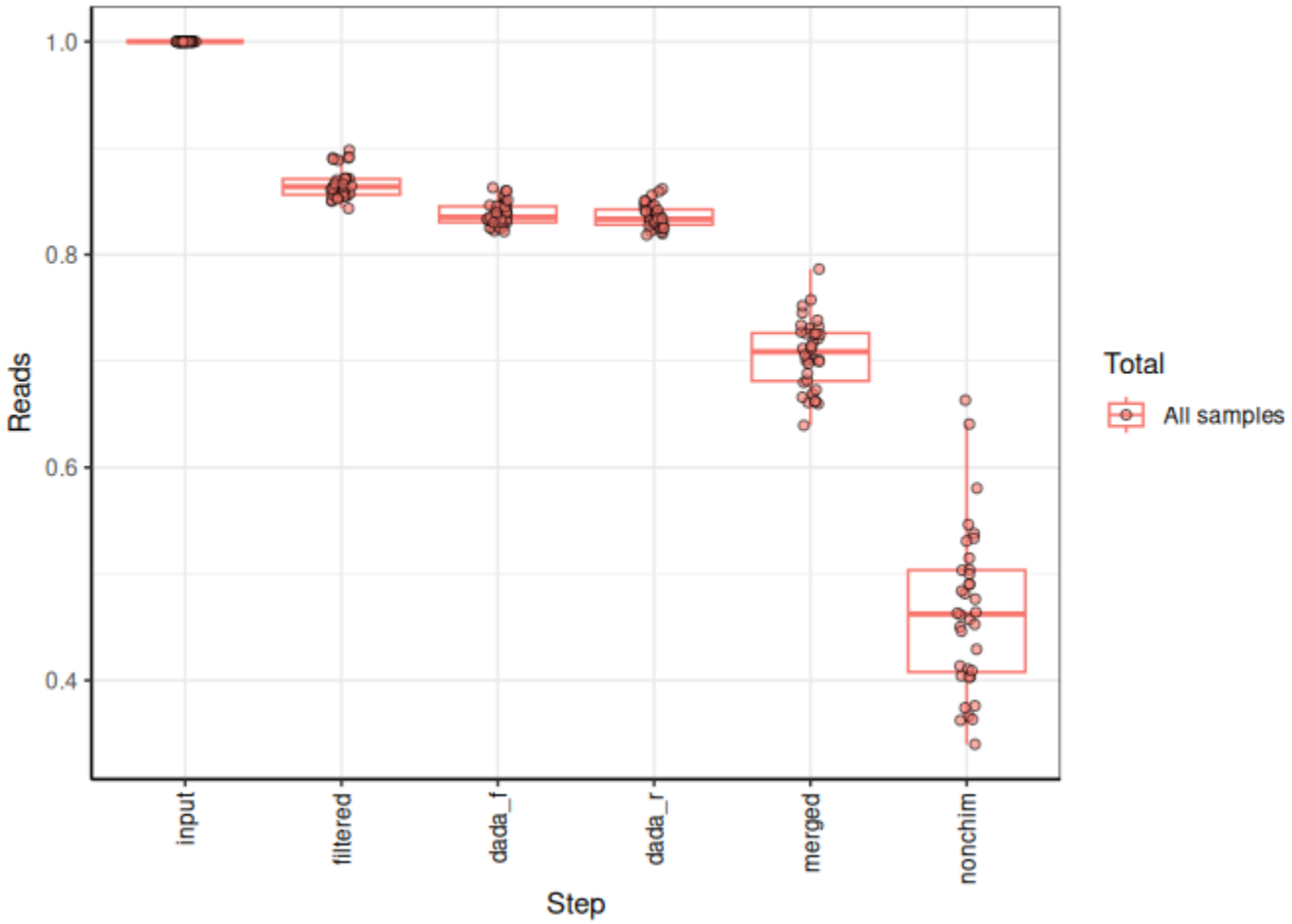
```
## AT50-12-TT-CTD4-1026m-DM                    42.9
## AT50-12-TT-CTD4-541m-DM                     48.2
## AT50-12-TT-CTD4-940m-DM                     37.4
## AT50-12-TT-CTD4-Bottom-DM                   58.1
## AT50-12-TT-CTD4-Surface-DM                  47.6
## AT50-12-TT-CTD6-Bottom-SM                   51.5
## AT50-12-TT-CTD7-WP1-SM                      53.1
## AT50-12-TT-CTD7-WP10-SM                     49.0
## AT50-12-TT-CTD7-WP4-SM                      50.4
## AT50-12-TT-CTD7-WP6-SM                      50.0
## AT50-12-TT-CTD7-WP8-SM                      49.0
## AT50-12-TT-CTD8-200m-SM                     36.3
## AT50-12-TT-CTD8-580m-SM                     40.3
## AT50-12-TT-CTD8-756m-SM                     40.9
## AT50-12-TT-CTD8-Bottom-SM                   48.4
## AT50-12-TT-CTD8-Surface-SM                  64.1
```

```
ggplot(melt(summary_tab[,-ncol(summary_tab)], variable.name='Step', value.name='Reads'),
aes(x = Step, y = Reads)) +
  geom_boxplot(outlier.shape = NA) + geom_jitter(aes(color=samp), height=0, width=0.2) +
  guides(fill='none') +
  theme_bw() + theme(axis.text.x = element_text(angle = 90, color='black', hjust=1, vjust=0.5),
                     axis.line.x = element_line(), axis.line.y = element_line())
```




```
#ggsave("read-count-tracking.pdf", width=8, height=6)

melt(summary_tab[,-ncol(summary_tab)], variable.name='Step', value.name='Reads') %>%
group_by(samp) %>%
  mutate(Reads=Reads/Reads[Step == "input"]) %>% mutate(Total='All samples') %>%
  ggplot(aes(x = Step, y = Reads)) +
```

```
  geom_boxplot(aes(color = Total), outlier.shape = NA) + theme_bw() +
  geom_point(position=position_jitterdodge(jitter.height=0, jitter.width=0.2),
  aes(fill=Total), alpha=0.6, pch=21) +
    theme(axis.text.x = element_text(angle = 90, color='black', hjust=1, vjust=0.5),
    panel.background = element_blank(),
          axis.line.x = element_line(), axis.line.y = element_line())
```




```
#ggsave("read-count-tracking_percent.pdf", width=8, height=6)
```

Annotation

```
library(DECIPHER)

load("/export/data1/db/16S_tag_processing_db/CABO-16S_20250911.RData")  # CABO-16S database, based on SILVA

## creating DNAStringSet object of our ASVs
dna <- DNAStringSet(getSequences(seqtab.nochim))

## and classify
tax_info <- IdTaxa(test=dna, trainingSet=trainingSet, strand="top", threshold=40, processors = 16)
```

Save the data as counts table and as fasta of ASV representative seqs

```
 # giving our seq headers more manageable names (ASV_1, ASV_2...)
asv_seqs <- colnames(seqtab.nochim)
asv_headers <- vector(dim(seqtab.nochim)[2], mode="character")

for (i in 1:dim(seqtab.nochim)[2]) {
  asv_headers[i] <- paste(">ASV", i, sep="_")
}
```

```
  # making and writing out a fasta of our final ASV seqs:
asv_fasta <- c(rbind(asv_headers, asv_seqs))
write(asv_fasta, "ASVs.fa")

  # count table:
asv_tab <- t(seqtab.nochim)
row.names(asv_tab) <- sub(">", "", asv_headers)
write.table(asv_tab, "ASVs_counts.tsv", sep="\t", quote=F, col.names=NA)

  # tax table:
  # creating table of taxonomy and setting any that are unclassified as "NA"
ranks <- c("domain", "phylum", "class", "order", "family", "genus", "species")
asv_tax <- lapply(tax_info, function(x) {
  taxa <- unlist(strsplit(x$taxon, ";"))[-1]  # drop "Root"
  taxa <- c(taxa, rep(taxa[length(taxa)], max(0,length(ranks)-length(taxa))))
  taxa <- taxa[1:length(ranks)]  # pesky euks
})
asv_tax <- matrix(unlist(asv_tax), ncol=length(ranks), byrow = T, dimnames = list(names(asv_tax), ranks))
colnames(asv_tax) <- ranks
rownames(asv_tax) <- gsub(pattern=">", replacement="", x=asv_headers)

write.table(asv_tax, "ASVs_taxonomy.tsv", sep = "\t", quote=F, col.names=NA)
```