# Supplementary Information

## Supplementary Methods, Figures S1-S8, Tables S1-S3

## for

# Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean

Yanmei Shi[1], Gene W. Tyson[1,3], John M. Eppley[1], Edward F. DeLong[1,2,*]

[1]Departments of Civil and Environmental Engineering and [2]Biological Engineering, Massachusetts Institute of Technology, Cambridge MA 02139

[3]Advanced Water Management Centre, University of Queensland, Brisbane, Queensland, Australia

*Corresponding Author

## Supplementary Methods

### Sample Collection and DNA/RNA extraction

Bacterioplankton samples (size fraction 0.22 µm – 1.6 mm) from the photic zone (25m, 75m, 125m) and the mesopelagic zone (500m) were collected from the Hawaii Ocean Time-series (HOT) Station ALOHA site in March 2006, as described previously (Shi et al 2009). Briefly, four replicate 1-liter seawater samples were prefiltered through 1.6-mm GF/A filters (Whatman, Maidstone, U.K.) and then filtered onto 0.22-µm Durapore filters (25mm diameter, Millipore, Bedford, MA) using a four-head peristaltic pump system. Each Durapore filter was immediately transferred to screw-cap tubes containing 1 ml of RNAlater (Ambion Inc., Austin, TX), and frozen at -80°C aboard the R/V Kilo Moana. Samples were transported frozen to the laboratory in a dry shipper and stored at -80°C until RNA extraction. Total sampling time, from arrival on deck to fixation in RNAlater was less than 20 minutes.

Replicate filters were pooled for RNA extractions, which were performed as previously described (Shi et al 2009), using the *mir*Vana™ RNA isolation kit (Ambion, Austin, TX). Samples were thawed on ice, and the 1 ml RNAlater was loaded onto two Microcon YM-50 columns (Millipore, Bedford, MA) to concentrate and desalt each sample. The resulting 50 µl of RNAlater was added back to the sample tubes, and total RNA extraction was performed following the *mir*Vana™ manual. Genomic DNA was removed using a Turbo DNA-free™ kit (Ambion, Austin, TX). Finally, extracted RNA (DNase-treated) from four replicate filters were combined, purified, and concentrated by using the MinElute PCR Purification Kit (Qiagen, Valencia, CA).

Bacterioplankton sampling for DNA extraction and DNA extraction was performed as previously described (Frias-Lopez et al 2008).

## RNA amplification and cDNA synthesis

Roughly 100 ng of total RNA was amplified using the MessageAmp II-Bacteria kit (Ambion) as described previously (Frias-Lopez et al 2008, Shi et al 2009). Briefly, total RNA was polyadenylated using Escherichia coli poly(A) polymerase. Polyadenylated RNA was converted to double-stranded cDNA via reverse transcription primed with an oligo(dT) primer containing a promoter sequence for T7 RNA polymerase and a recognition site for the restriction enzyme BpmI (T7-BpmI-(dT)$^{16}$VN, GCCAGTGAATTG**TAATACGACTCACTATA**GGGGCGACTGGAGTTTTTTTTT TTTTTTVN). cDNA was then transcribed in vitro at 37 °C for 6 hours, yielding large quantities (~100 ug) of antisense RNA. An aliquot of antisense RNA (~5 ug aliquot) was polyadenylated again and converted to double-stranded cDNA using first the SuperScript III First-Strand Synthesis System (Invitrogen, Carlsbad, CA, USA) with priming via oligo(dT) for first-strand synthesis, and then the SuperScript Double-Stranded cDNA synthesis kit (Invitrogen) for second-strand synthesis. cDNA was then purified with the QIAquick PCR purification kit (Qiagen), digested with BpmI for 2-3 hours at 37 °C to remove poly(A) tails, and purified again with the QIAquick PCR purification kit. Purified cDNA was used for the generation of single-stranded DNA libraries and emulsion PCR according to established protocols (454 Life Sciences, Roche). Clonally amplified library fragments were then sequenced on a Genome Sequencer GS20 System (Roche).

## Bioinformatics analyses

Taxonomic classification of 16S rRNA sequences. Ribosomal RNA sequences were first identified by comparing the data sets to a combined 5S, 16S, 18S, 23S, and 28S rRNA database derived from available microbial genomes and sequences from the ARB SILVA LSU and SSU databases (www.arb-silva.de). 16S rRNA sequences were then selected by BLASTing (Altschul et al 1990) against SILVA SSU databases (bits score ≥ 50, alignment length ≥ 80% of the read length, and alignment length ≥ 100bp), and classified using the online Greengenes classifier tools (http://greengenes.lbl.gov/cgi-bin/nph-classify.cgi), using the Hugenholtz taxonomy. The parameters used for classifying 16S rRNA were a minimum alignment length of 100bp, and a minimum sequence identity of 75%. For the shotgun sequences, 16S rRNA reads were chosen based on the cutoff of a bits score ≥ 50 and an alignment length ≥ 280bp, and the parameters used for classifying 16S rRNA were a minimum alignment length of 280bp, and a minimum sequence identity of 75%.

Taxonomic classification of protein-coding sequences. Protein-coding sequences were identified by blasting against the NCBI non-redundant (NCBI-nr) protein database. The BLASTx output was parsed to analyze the taxonomic breakdown using MEGAN (Huson et al 2007), with bit scores > 40 within 10% of the top scoring hits.

Functional analyses using the SEED database and GOS protein cluster database. Non rRNA reads were assigned to SEED subsystems and GOS protein clusters based on top BLASTx hits with bits score ≥ 40. A bootstrapping method (Rodriguez-Brito et al 2006), which takes care of the size difference among subsystems and looks for statistically significant differences metagenomes, was applied to identify subsystems that were enriched in the cDNA libraries relative to the corresponding DNA libraries. GOS protein cluster-based analysis was perform as previously described (Frias-Lopez et al 2008). Briefly, cluster-based expression ratios were calculated as the number of reads found for each protein cluster in the cDNA library relative to that found in the DNA library, which was further normalized for the difference in DNA and cDNA library size. Functional annotations for GOS protein clusters, when available, were available from a study by Yooseph *et al* (Yooseph et al 2007). The cluster-based expression ratios were ranked from highest to lowest (Figure 3) to look at clusters being expressed at elevated levels.

Reference genome-centric analysis. Two custom databases (one nucleotide database and one amino acid database) were constructed from 2067 publicly available microbial genome sequences and annotations (fully sequenced and draft genomes as of January 2009). Non-rRNA cDNA and DNA reads from all four depths were compared against the custom nucleotide database, and reads with top hit bits score $\geq 40$ were assigned to the corresponding genome. In order to compensate for likely uneven phylogenetic representation in the databases, we allowed any read to map to several reference read with the same alignment score. Recruitment of protein-coding cDNA and DNA reads onto reference genomes were performed by assigning reads to top amino acid sequences with bits score $\geq 40$. For each ORF, recruited cDNA abundance was divided by the recruited DNA abundance, to give an indication of per-copy cDNA level. If there were cDNA hits but no DNA hits for a given ORF, the number of DNA hits was considered as 1.

To examine the expression of *Pelagibacter* strain HTCC7211-specific ORFs, putative *Pelagibacter* reads were first identified as reads with top BLASTx hit (against NCBI-nr) to *Pelagibacter* and with a bit score >40. Each of these putative *Pelagibacter* reads then was searched against a custom database of *Pelagibacter* ORFs derived from 3 fully sequenced *Pelagibacter* strains (HTCC1062, HTCC1002, HTCC7211) using BLASTx, and assigned to the best hit ORF. The HTCC7211-specific ORFs were identified as ORFs with no best reciprocal hit, based on the cutoff of a minimum sequence identity of 30%, and a minimum alignment length fraction of 75%, in the genomes of HTCC1062 or HTCC1002.

## Figure Legends

**Figure S1. Biogeochemical data of the sampling station collected on the cruise.** Dashed lines indicate four sampling depths. Data source: http://hahana.soest.hawaii.edu/hot/hot-dogs/interface.html.

**Figure S2. Taxonomic classification based on 16S rRNA-bearing shotgun sequences.** The shotgun libraries and pyrosequencing libraries were constructed from identical DNA samples. Taxonomic assignments were binned at the Order level, using the Hugenholtz taxonomy of Greengenes (see Supplementary Methods). 16S rRNA sequences that could not be classified were excluded from the analysis. Y-axis scale represents the percentage of the total classified 16S rRNA reads. Only taxa that

represented ≥ 1% of all classified reads are displayed.

**Figure S3. Stacked area plot showing taxonomic classification of protein-coding sequences.** Taxonomic assignments were based on BLASTx against NCBI-nr protein database, using MEGAN (Huson et al 2007), with default settings. Upper panel represents DNA samples, and lower panel represents cDNA samples.

**Figure S4. Abundance and normalized expression levels of genes involved in nitrogen metabolism.** The abundance of 16S rRNA genes was used to indicate taxon abundance, and was compared to detected abundance of a suite of functional genes (listed in figure legends). Normalized gene expression was calculated as described in Supplementary Methods. (A) Functional genes putatively originated from *Prochlorococcus* populations, in the three euphotic zone samples. (B) Functional genes putatively originated from marine group I crenarchaeota populations in the deep euphotic zone and the mesoplegic samples.

**Figure S5. Abundance, expression and taxonomic origins of Proteorhodopsin (PR)-encoding reads.** (A). Representation of PR-encoding reads in the DNA and cDNA data sets, and their normalized expression levels in the four depths. (B) Putative taxonomic breakdown of PR sequence reads. PR sequences were first identified by BLASTx against NCBI-nr database, then aligned to a custom PR sequence database (McCarren and DeLong 2007), and finally added to the backbone PR phylogenetic tree

using ARB's “parsimony insertion” feature. The taxonomic origin of a PR-encoding sequence was assumed the same as that of the most related sequence in the PR phylogenetic tree.

**Figure S6. Expression of genes involved in aerobic anoxygenic phototrophy (AAP), using a *Roseobacter*-like BAC clone insert as a reference.** The BAC clone is eBACred25D05 with an accession number of AY671989. *puf*: light-harvesting and reaction center genes; *bch*: bacteriochlorophyll biosynthesis genes; *crt*, carotenoid biosynthesis genes. Y-axis scale represents normalized cDNA to DNA ratio (normalized expression level; see Supplementary Methods).

**Figure S7. Gene expression of *Pelagibacter* HTCC7211-specific ORFs.** The HTCC7211-specific ORFs are denoted by the black dots on top the panel, and were identified as ORFs lack of apparent homology to ORFs in the two coastal *Pelagibacter* strains HTCC1062 and HTCC1002 (see Supplementary Methods).

**Figure S8. Genome-wide expression profiles of marine crenarchaea-related populations, in all four depths.** The x-axis, y-axis, and figure legend are the same as those in Figure 5.

Supplementary information is available at the ISME journal’s website.

# REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic Local Alignment Search Tool. *J Mol Biol* **215:** 403-410.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al* (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* **105:** 3805-3810.

Huson DH, Auch AF, Qi J, Schuster SC (2007). MEGAN analysis of metagenomic data. *Genome Res* **17:** 377-386.

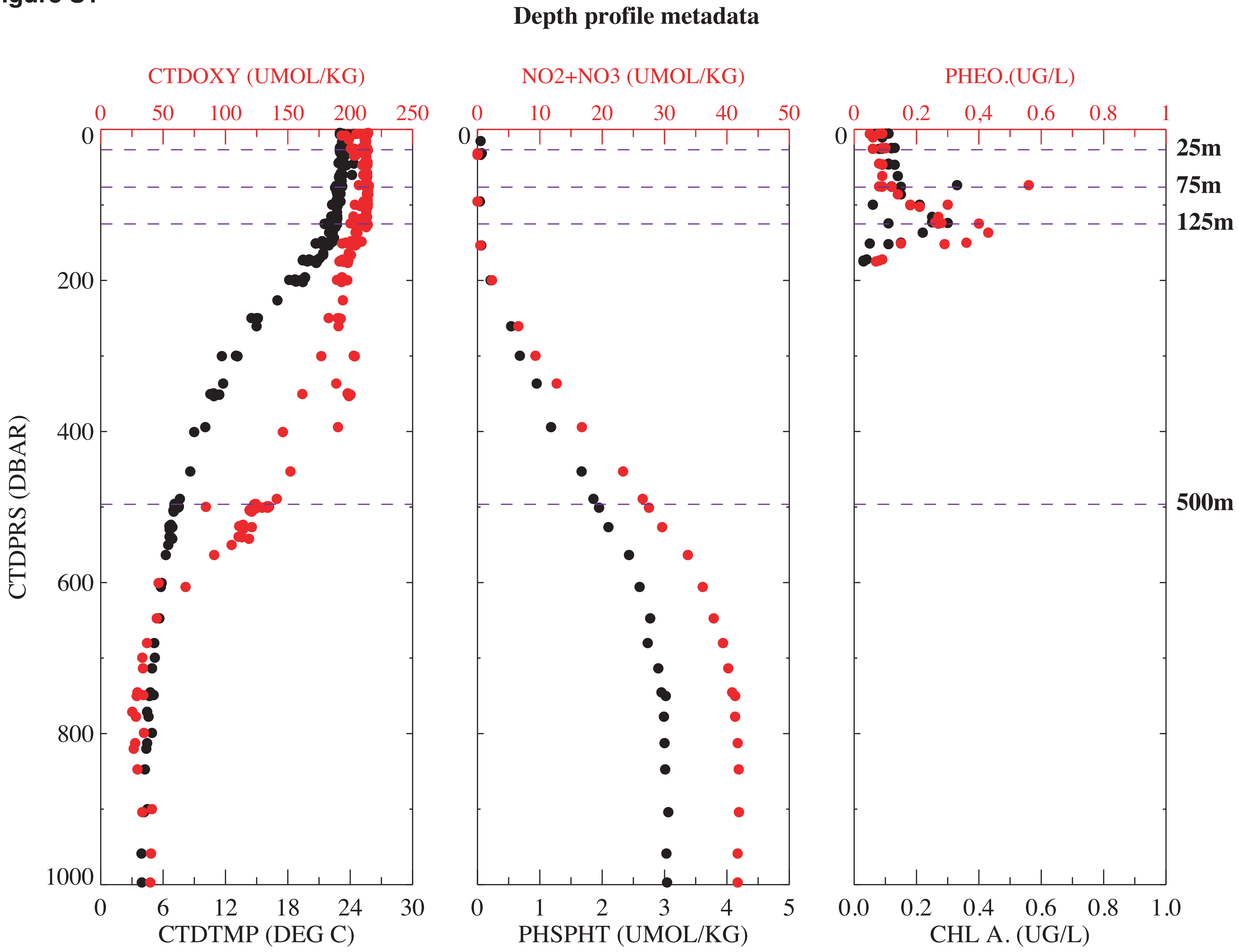
McCarren J, DeLong EF (2007). Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environ Microbiol* **9:** 846-858.

Rodriguez-Brito B, Rohwer F, Edwards RA (2006). An application of statistics to comparative metagenomics. *BMC Bioinformatics* **7:** doi:10.1186/1471-2105-1187-1162.

Shi Y, Tyson GW, DeLong EF (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459:** 266-269.

Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K *et al* (2007). The Sorcerer II Global Ocean Sampling expedition: Expanding the universe of protein families. *PLoS Biol* **5:** 432-466.

**Figure S1**


Depth profile metadata
CTDOXY (UMOL/KG)
0
50
100
150
200
250
NO2+NO3 (UMOL/KG)
0
10
20
30
40
50
PHEO.(UG/L)
0
0.2
0,4
0.6
0.8
1
CTDPRS (DBAR)
0
200
400
600
800
1000
25m
75m
125m
500m
0
6
12
18
24
30
CTDTMP (DEG C)
0
1
2
3
4
5
PHSPHT (UMOL/KG)
0.0
0.2
0.4
0.6
0.8
1.0
CHL A. (UG/L)

# Figure S2

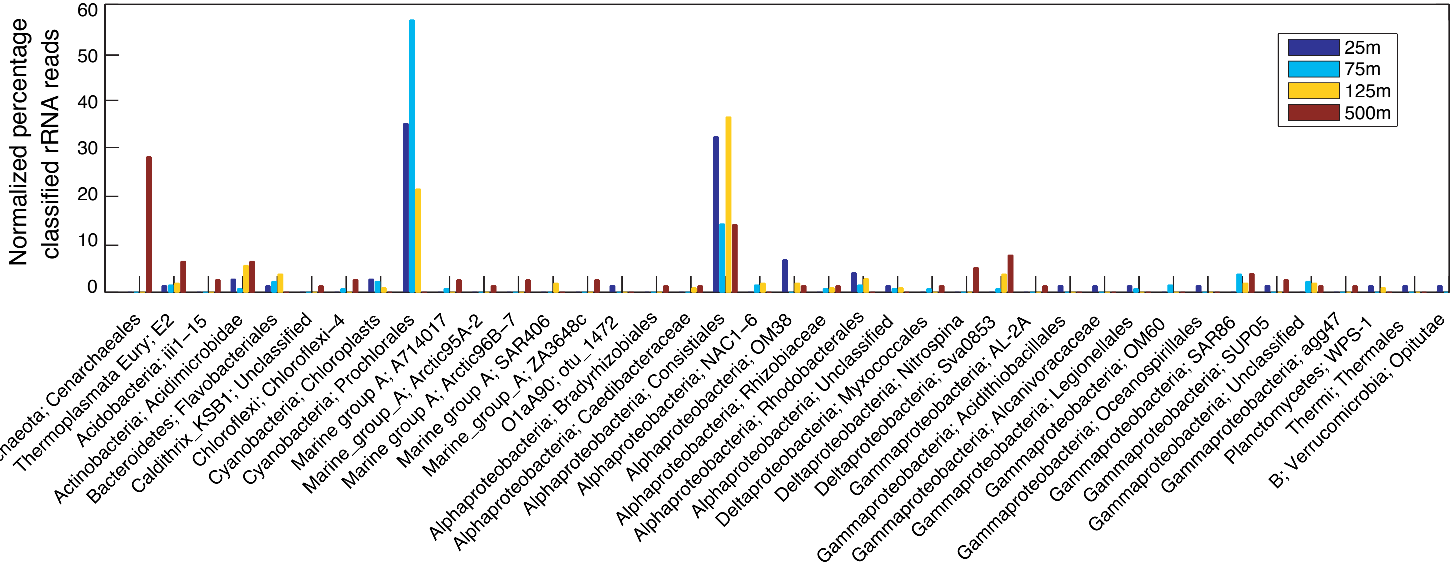
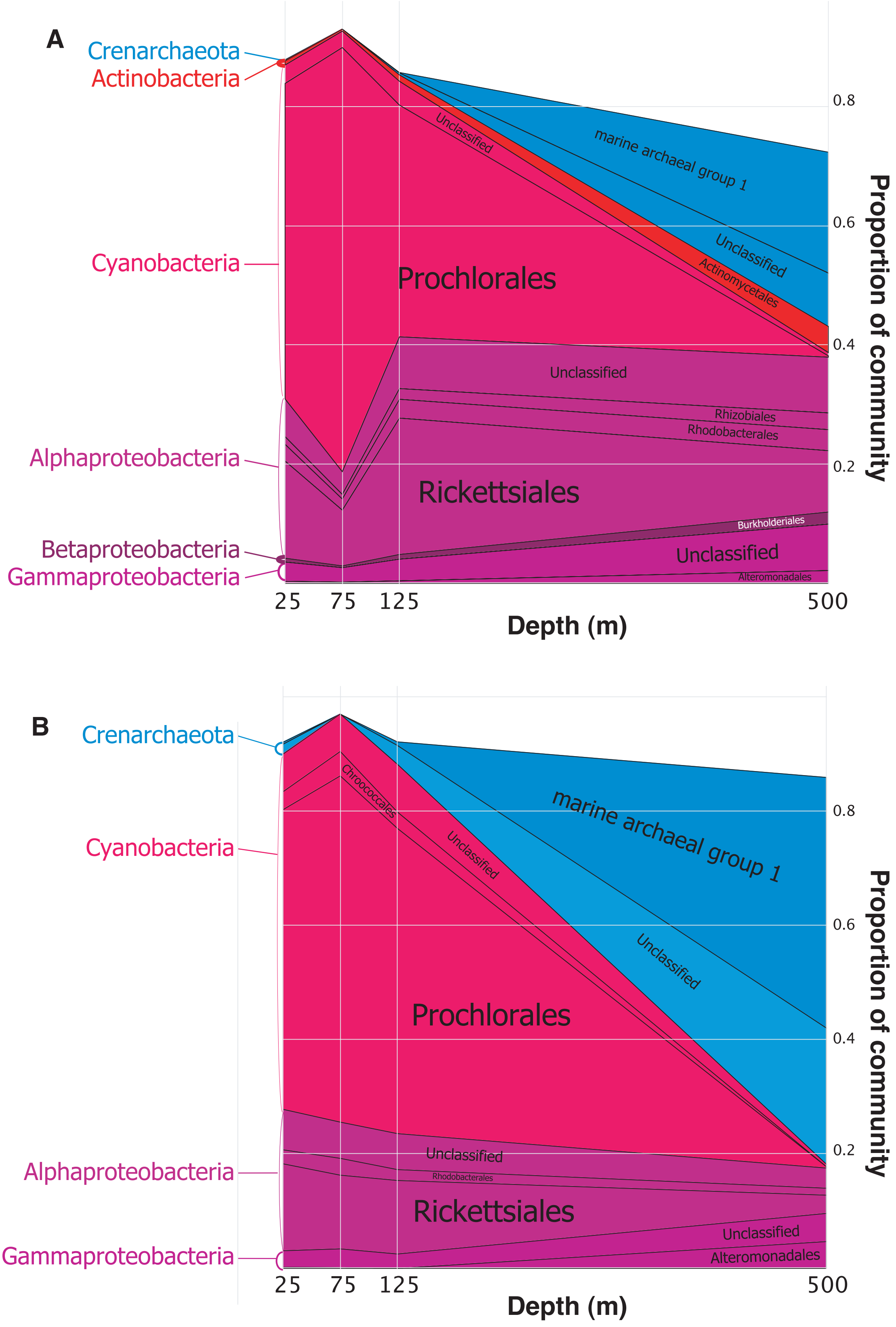
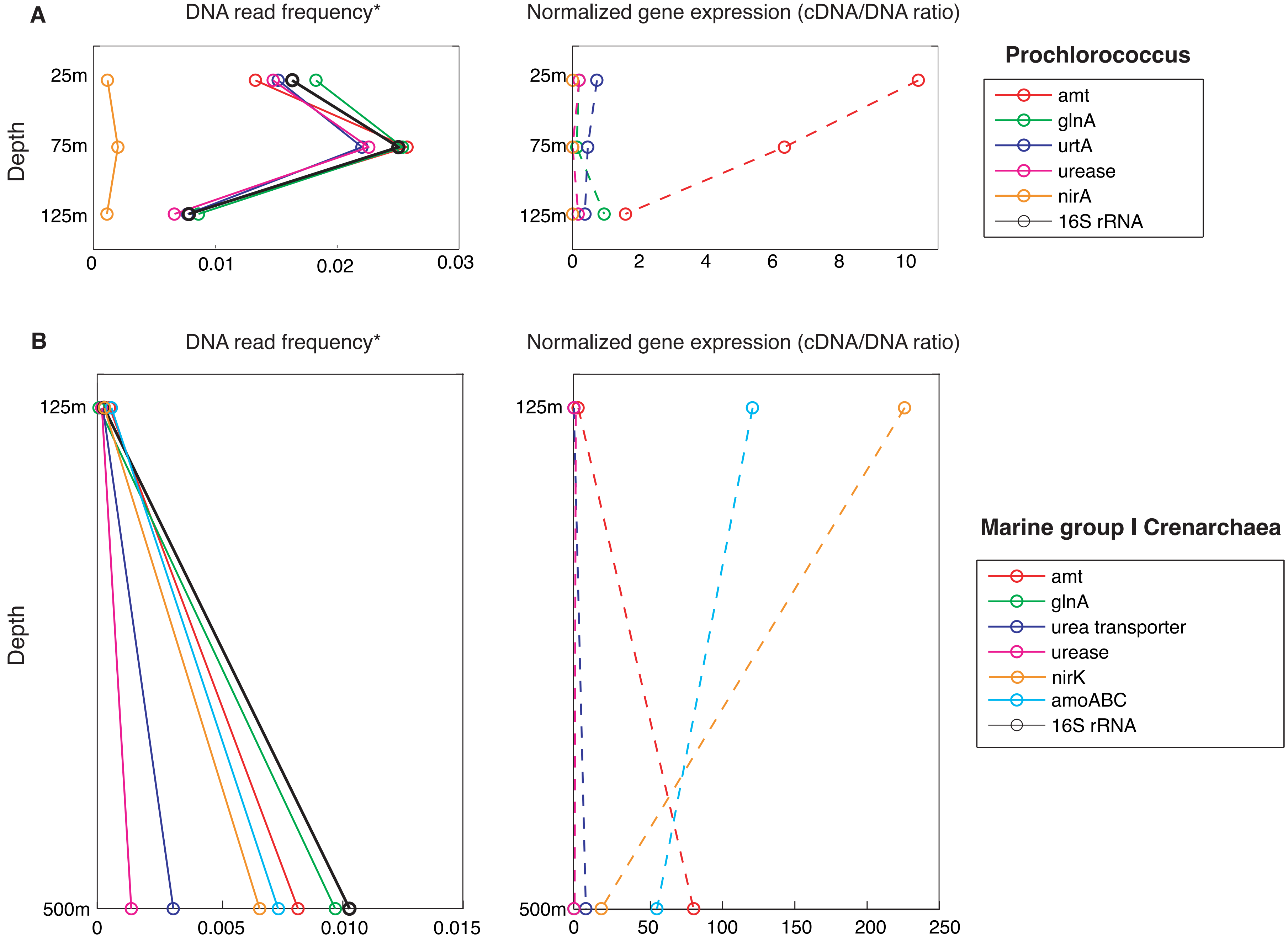

Normalized percentage
classified rRNA reads
60
50
40
30
20
10
0
25m
75m
125m
500m
Thaumarchaeota; Cenarchaeales
Thermoplasmata Eury; E2
Acidobacteria; iii1–15
Actinobacteria; Acidimicrobidae
Bacteroidetes; Flavobacteriales
Caldithrix_KSB1; Unclassified
Chloroflexi; Chloroflexi–4
Cyanobacteria; Chloroplasts
Cyanobacteria; Prochlorales
Marine group A; A714017
Marine_group_A; Arctic95A-2
Marine group A; Arctic96B–7
Marine group A; SAR406
Marine_group_A; ZA3648c
O1aA90; otu_1472
Alphaproteobacteria; Bradyrhizobiales
Alphaproteobacteria; Caedibacteraceae
Alphaproteobacteria; Consistiales
Alphaproteobacteria; NAC1–6
Alphaproteobacteria; OM38
Alphaproteobacteria; Rhizobiaceae
Alphaproteobacteria; Rhodobacterales
Alphaproteobacteria; Unclassified
Deltaproteobacteria; Myxococcales
Deltaproteobacteria; Nitrospina
Deltaproteobacteria; Sva0853
Gammaproteobacteria; AL-2A
Gammaproteobacteria; Acidithiobacillales
Gammaproteobacteria; Alcanivoracaceae
Gammaproteobacteria; Legionellales
Gammaproteobacteria; OM60
Gammaproteobacteria; Oceanospirillales
Gammaproteobacteria; SAR86
Gammaproteobacteria; SUP05
Gammaproteobacteria; Unclassified
Gammaproteobacteria; agg47
Planctomycetes; WPS-1
Thermi; Thermales
B; Verrucomicrobia; Opitutae

Figure S3

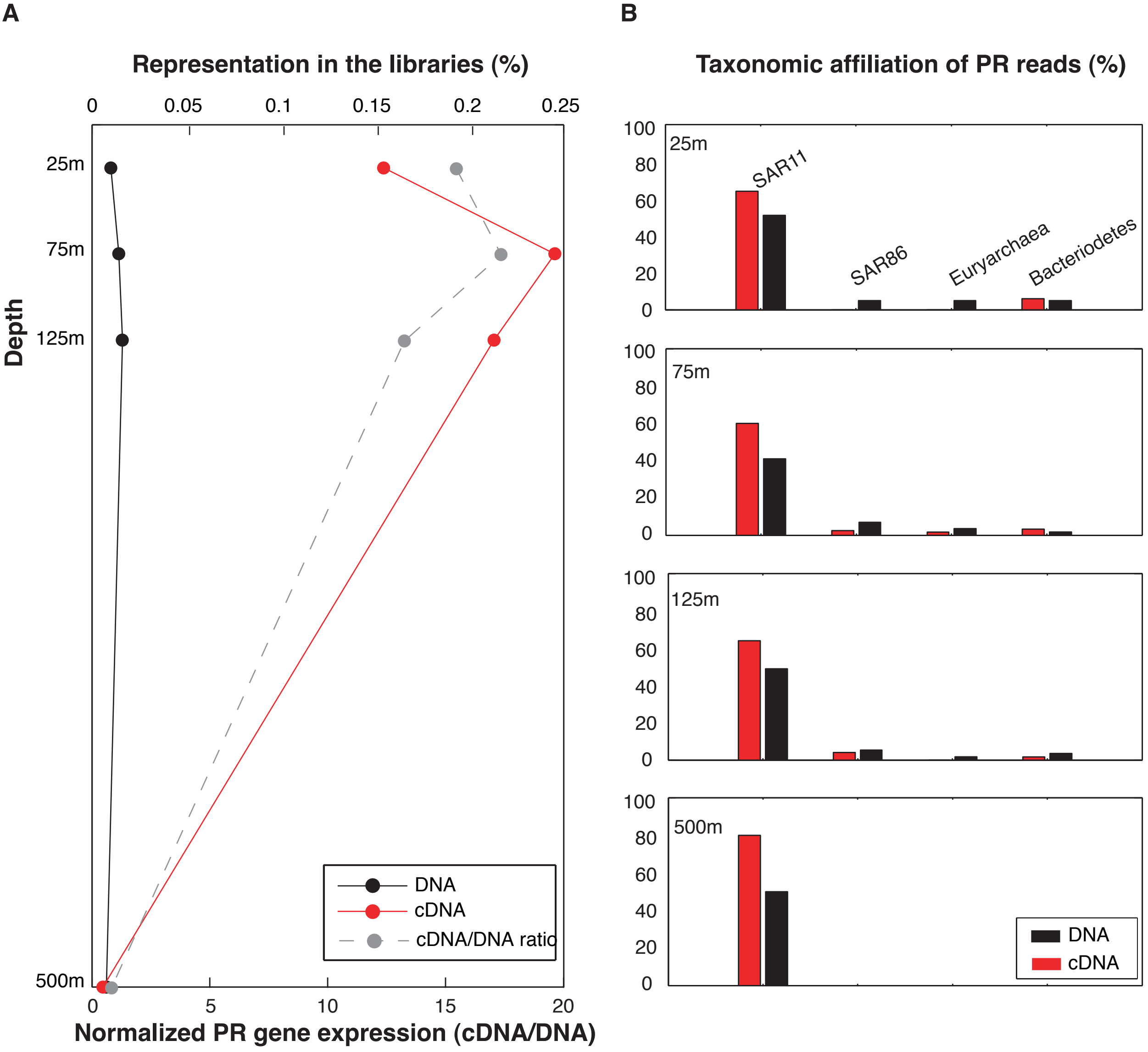

A
Crenarchaeota
Actinobacteria
Cyanobacteria
Alphaproteobacteria
Betaproteobacteria
Gammaproteobacteria
marine archaeal group 1
Unclassified
Actinomycetales
Unclassified
Prochlorales
Unclassified
Rhizobiales
Rhodobacterales
Rickettsiales
Burkholderiales
Unclassified
Alteromonadales
0.8
0.6
0.4
0.2
Proportion of community
25
75
125
500
Depth (m)
B
Crenarchaeota
Cyanobacteria
Alphaproteobacteria
Gammaproteobacteria
marine archaeal group 1
Unclassified
Chroococcales
Unclassified
Prochlorales
Unclassified
Rhodobacterales
Rickettsiales
Unclassified
Alteromonadales
0.8
0.6
0.4
0.2
Proportion of community
25
75
125
500
Depth (m)

## Figure S4

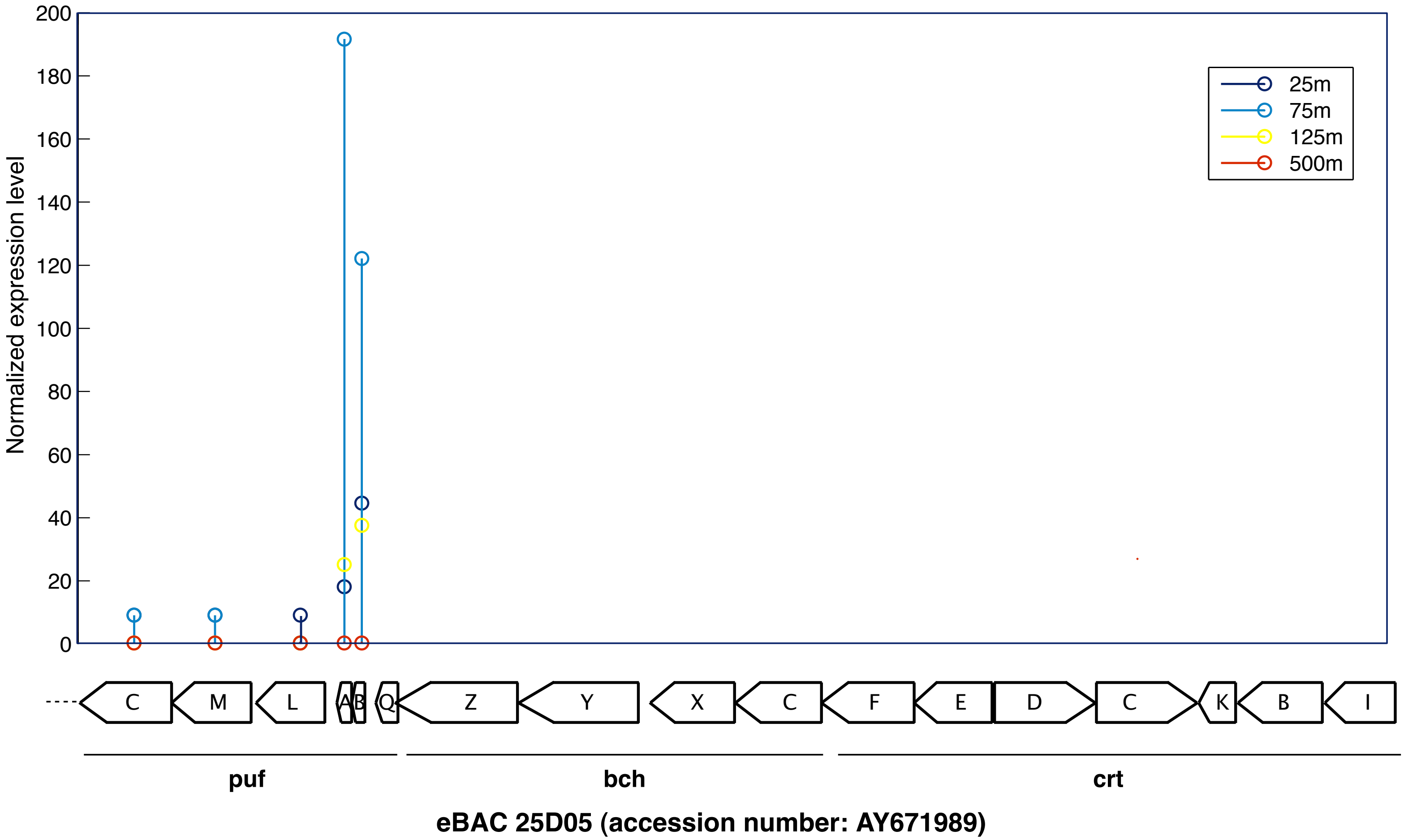


* The percentage is normalized to the gene length (per kb)

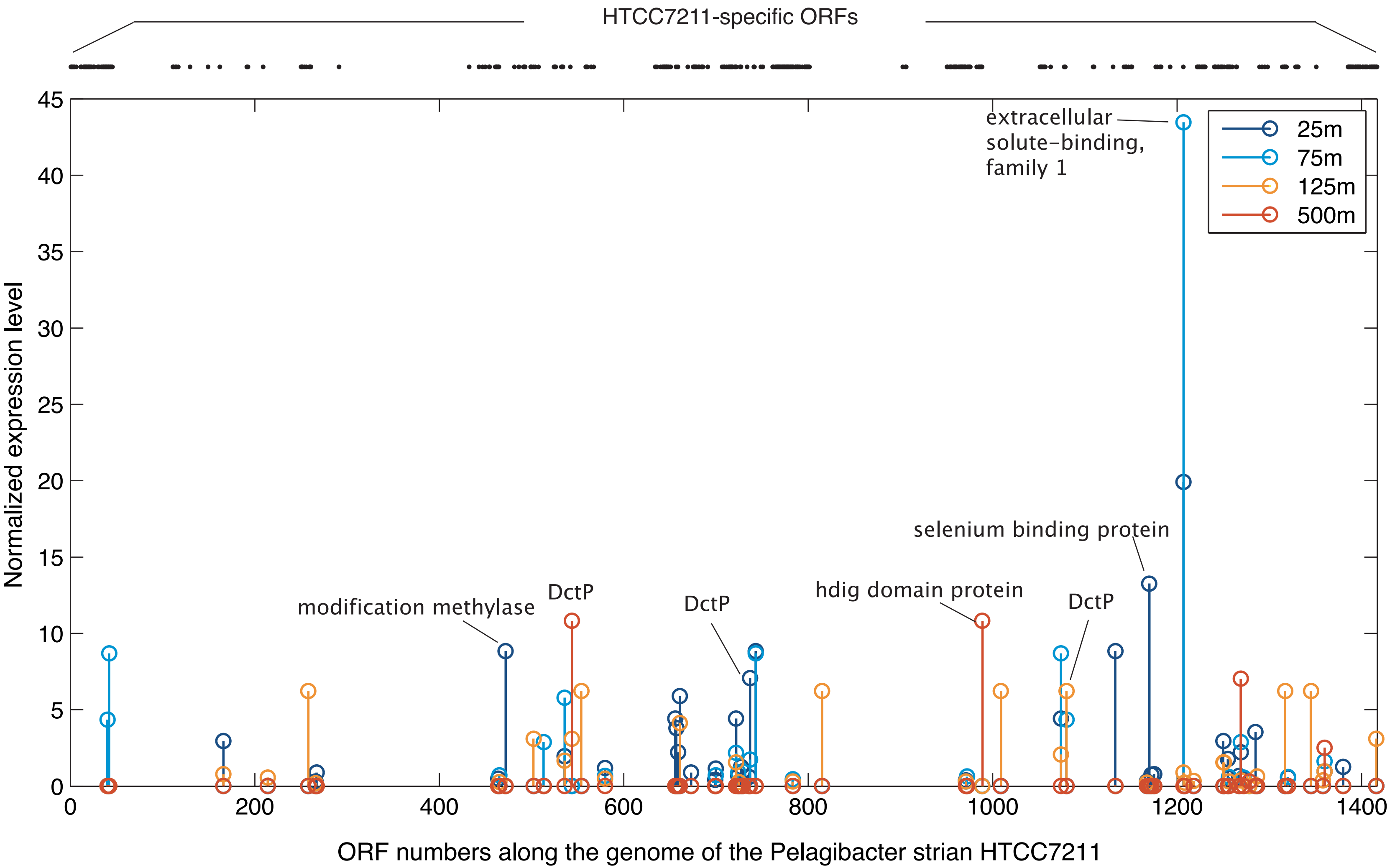
Figure S5


A
Representation in the libraries (%)
0
0.05
0.1
0.15
0.2
0.25
Depth
25m
75m
125m
500m
DNA
cDNA
cDNA/DNA ratio
0
5
10
15
20
Normalized PR gene expression (cDNA/DNA)
B
Taxonomic affiliation of PR reads (%)
100
80
60
40
20
0
25m
SAR11
SAR86
Euryarchaea
Bacteriodetes
75m
125m
500m
DNA
cDNA

## Figure S6


200
180
160
140
120
100
80
60
40
20
0
Normalized expression level
25m
75m
125m
500m
C
M
L
A
B
Q
Z
Y
X
C
F
E
D
C
K
B
I
puf
bch
crt
eBAC 25D05 (accession number: AY671989)

# Figure S7


HTCC7211-specific ORFs
45
40
35
30
25
20
15
10
5
0
Normalized expression level
0
200
400
600
800
1000
1200
1400
ORF numbers along the genome of the Pelagibacter strian HTCC7211
25m
75m
125m
500m
extracellular solute-binding, family 1
selenium binding protein
hdig domain protein
DctP
DctP
DctP
modification methylase

## Figure S8

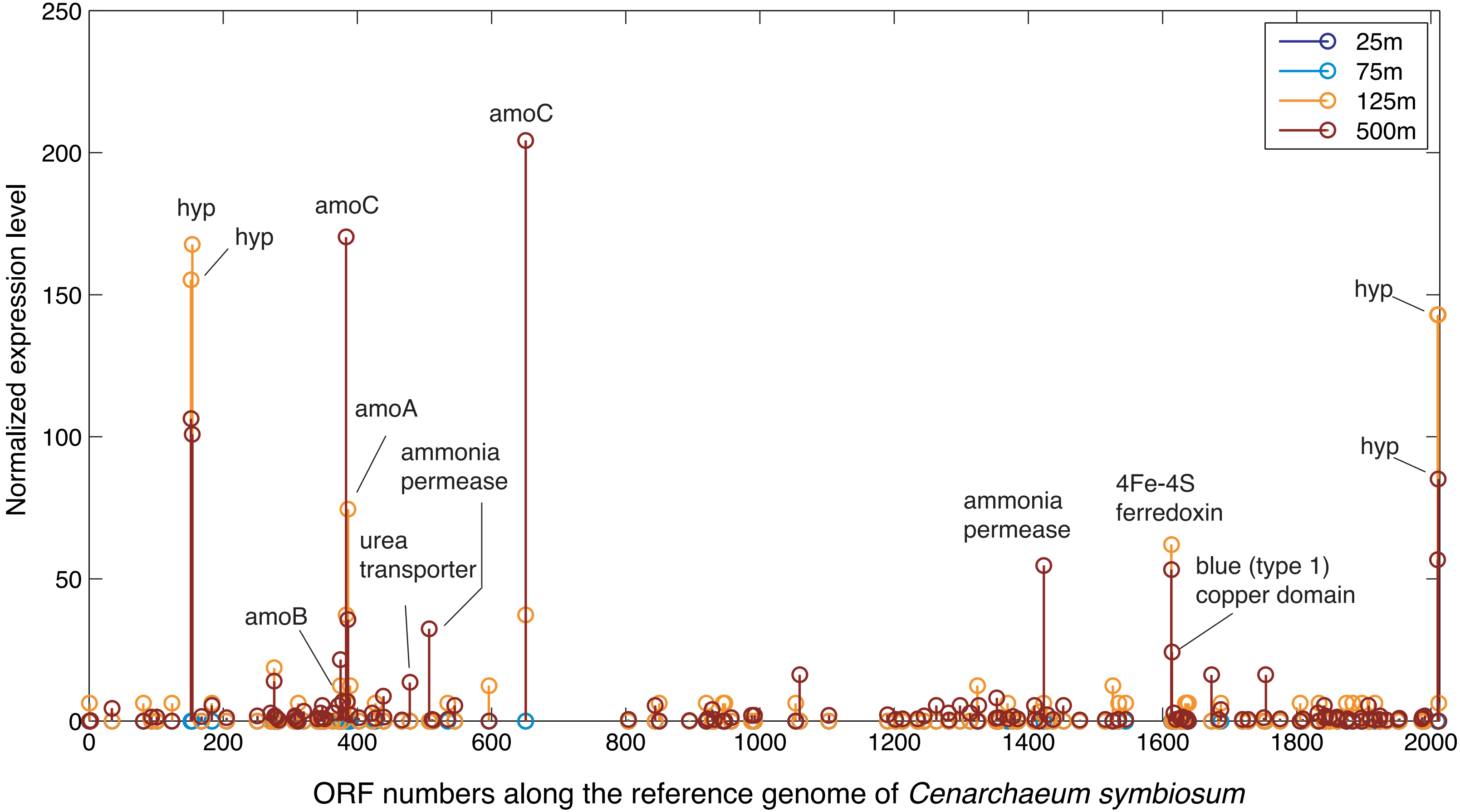

Normalized expression level
250
200
150
100
50
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000
ORF numbers along the reference genome of *Cenarchaeum symbiosum*
25m
75m
125m
500m
amoC
hyp
hyp
amoC
amoA
ammonia permease
urea transporter
amoB
hyp
hyp
ammonia permease
4Fe-4S ferredoxin
blue (type 1) copper domain

**Table S1. Comparison of Prochlorales representation in HF (DeLong et al, 2006) and HOT 179 fosmid clone libraries.**

| Fosmid library | Sampling depth | # reads assigned to a taxon* | # (%) reads assigned to Prochlorales |
|---|---|---|---|
| **HF** | 10 m | 5165 | 341 (6.6%) |
| | 75 m | 5953 | 124 (2.1%) |
| | 130 m | 4530 | 169 (3.7%) |
| | 500 m | 6777 | 6 (0.09%) |
| **HOT179** | 25 m | 8196 | 820 (10%) |
| | 75 m | 10120 | 1502 (14.8%) |
| | 125 m | 15375 | 1300 (8.5%) |
| | 500 m | 16544 | 22 (0.13%) |

* Taxon breakdown was performed with MEGAN (Huson *et al*, 2007), using the following LCA parameters: min support = 1, min score = 70, top percent = 0.

## Table S2. Recruitment of cDNA and DNA reads to abundant reference genomes.

| Reference genomes | # of DNA reads assigned to a reference genome | | | | # of cDNA reads assigned to a reference genome | | | |
|---|---|---|---|---|---|---|---|---|
| | 25m | 75m | 125m | 500m | 25m | 75m | 125m | 500m |
| Prochlorococcus marinus AS9601 | 28682 | 43034 | 10311 | 23 | 1656 | 1900 | 1926 | 4 |
| Prochlorococcus marinus MIT 9301 | 24272 | 37042 | 8733 | 19 | 1683 | 2081 | 1887 | 7 |
| Prochlorococcus marinus MIT 9312 | 14405 | 22578 | 5805 | 12 | 926 | 1125 | 1043 | 2 |
| Prochlorococcus marinus MIT 9215 | 14354 | 21886 | 5193 | 21 | 5039 | 1902 | 2275 | 18 |
| Prochlorococcus marinus MED4 | 1277 | 2737 | 644 | 5 | 197 | 269 | 163 | 0 |
| Candidatus Pelagibacter ubique HTCC1062 | 1137 | 1241 | 1642 | 612 | 238 | 204 | 291 | 84 |
| Candidatus Pelagibacter ubique B HTCC1002 | 1102 | 1242 | 1616 | 628 | 232 | 196 | 262 | 102 |
| Psychroflexus torquis ATCC 700755 ATCC700755 | 1383 | 1287 | 1436 | 181 | 170 | 195 | 187 | 30 |
| Prochlorococcus marinus NATL1A | 126 | 847 | 2571 | 2 | 13 | 43 | 569 | 0 |
| Prochlorococcus marinus NATL2A | 111 | 786 | 2511 | 5 | 15 | 51 | 595 | 2 |
| Synechococcus CC9605 | 1421 | 1485 | 335 | 2 | 64 | 80 | 54 | 2 |
| Prochlorococcus marinus MIT 9515 | 540 | 1042 | 243 | 4 | 59 | 86 | 120 | 0 |
| Synechococcus sp WH8102 | 146 | 272 | 35 | 0 | 16 | 29 | 16 | 1 |
| Alteromonas macleodii Deep ecotype | 10 | 2 | 2 | 426 | 4 | 3 | 5 | 406 |
| Prochlorococcus marinus phi P-SSM4 | 179 | 104 | 55 | 0 | 18 | 9 | 4 | 0 |
| Nitrosopumilus maritimus SCM1 | 1 | 2 | 44 | 260 | 0 | 2 | 188 | 1728 |
| Prochlorococcus marinus phi P-SSM2 | 135 | 74 | 51 | 0 | 4 | 3 | 1 | 0 |
| Prochlorococcus marinus CCMP1375 | 19 | 24 | 126 | 3 | 0 | 1 | 58 | 3 |
| OM42 clade HTCC2255 | 36 | 45 | 50 | 24 | 6 | 8 | 11 | 10 |
| Erythrobacter sp. SD-21 | 69 | 7 | 13 | 55 | 2 | 0 | 2 | 5 |
| Acinetobacter baumannii SDF | 101 | 9 | 2 | 27 | 0 | 0 | 0 | 2 |
| Prochlorococcus marinus str. MIT 9211 MIT9211 | 13 | 25 | 91 | 2 | 0 | 3 | 34 | 0 |
| Tenacibaculum sp. MED152 | 35 | 31 | 31 | 17 | 9 | 3 | 12 | 1 |
| Prochlorococcus marinus MIT9313 | 2 | 4 | 101 | 1 | 2 | 1 | 21 | 24 |
| Prochlorococcus marinus MIT 9303 | 9 | 6 | 87 | 0 | 1 | 0 | 12 | 2 |
| Synechococcus RCC307 | 37 | 31 | 14 | 8 | 1 | 4 | 12 | 0 |
| Synechococcus sp. RS9916 RS9917 | 28 | 35 | 15 | 2 | 6 | 2 | 12 | 0 |
| Flavobacteriales bacterium ALC-1 | 17 | 23 | 23 | 12 | 3 | 9 | 8 | 1 |
| Kordia algicida OT-1 | 22 | 27 | 14 | 11 | 5 | 2 | 3 | 0 |
| Acinetobacter baumannii ACICU | 44 | 2 | 11 | 11 | 0 | 0 | 0 | 1 |
| Rhodospirillales sp. BAL199 | 15 | 13 | 8 | 33 | 3 | 6 | 10 | 7 |
| Candidatus Vesicomyosocius okutanii HA | 2 | 4 | 8 | 54 | 0 | 3 | 5 | 7 |
| Pseudomonas syringae phaseolicola 1448A | 38 | 5 | 4 | 19 | 2 | 1 | 1 | 4 |
| Candidatus Ruthia magnifica | 7 | 4 | 3 | 51 | 1 | 2 | 6 | 78 |
| Xanthomonas campestris B100 | 36 | 6 | 3 | 15 | 0 | 0 | 0 | 15 |
| marine gamma proteobacterium HTCC2080 | 24 | 14 | 11 | 7 | 18 | 11 | 8 | 5 |
| Flavobacteriales sp. SCB49 | 25 | 13 | 9 | 5 | 2 | 5 | 3 | 3 |
| Flavobacteriales sp. BAL38 | 23 | 12 | 13 | 4 | 1 | 3 | 7 | 2 |
| Synechococcus sp. WH5701 | 14 | 14 | 13 | 11 | 2 | 4 | 7 | 3 |
| Staphylococcus aureus phi G1 | 45 | 4 | 1 | 1 | 0 | 0 | 0 | 0 |
| Brevundimonas sp. BAL3 | 27 | 6 | 2 | 16 | 1 | 0 | 2 | 0 |

**Table S3. Normalized gene expression of *Pelagibacter* strain HTCC7211 (top 60 highly expressed).**

| ORF number | 25m | 75m | 125m | 500m | annotation |
|---|---|---|---|---|---|
| 1207 | 19.9 | 43.3 | 0.9 | 0.0 | extracellular solute-binding protein, family 1 |
| 1263 | 3.2 | 14.9 | 1.8 | 37.8 | spermidine/putrescine-binding periplasmic protein |
| 507 | 22.1 | 28.6 | 18.6 | 1.8 | bacteriorhodopsin |
| 244 | 26.5 | 2.9 | 0.0 | 0.0 | protein of unknown function |
| 623 | 8.8 | 26.0 | 2.1 | 0.0 | conserved hypothetical protein |
| 631 | 22.7 | 8.7 | 9.9 | 0.0 | acetaldehyde dehydrogenase II (acdh-ii) |
| 1226 | 0.0 | 2.9 | 18.6 | 21.6 | conserved hypothetical protein |
| 664 | 0.0 | 0.0 | 18.6 | 0.0 | hypothetical protein |
| 1019 | 3.2 | 3.9 | 3.1 | 16.2 | ABC transporter |
| 1094 | 1.7 | 13.6 | 2.3 | 3.2 | Bacterial extracellular solute-binding protein, family 7 |
| 371 | 13.2 | 1.4 | 1.2 | 0.0 | heAt shock protein a |
| 1170 | 13.2 | 0.0 | 0.0 | 0.0 | selenium binding protein |
| 914 | 5.3 | 13.0 | 4.1 | 0.0 | ribosomal protein L13 |
| 1328 | 0.0 | 4.3 | 12.4 | 0.0 | pilin (bacterial filament) |
| 954 | 12.4 | 3.3 | 0.9 | 0.0 | conserved hypothetical protein |
| 1159 | 11.8 | 0.6 | 0.5 | 0.5 | GTP cyclohydrolase I |
| 243 | 10.1 | 11.0 | 3.2 | 5.4 | Na+/solute symporter (Ssf family) |
| 989 | 0.0 | 0.0 | 0.0 | 10.8 | hdig domain protein |
| 544 | 0.0 | 0.0 | 3.1 | 10.8 | trap dicarboxylate transporter, dctp subunit |
| 292 | 8.8 | 5.8 | 5.0 | 10.8 | trap dicarboxylate transporter - dctp subunit |
| 286 | 1.8 | 0.0 | 0.9 | 10.8 | conserved hypothetical protein |
| 195 | 0.5 | 0.0 | 0.0 | 10.8 | chaperone protein DnaJ |
| 823 | 0.0 | 1.1 | 9.3 | 0.0 | transcription termination/antitermination factor NusG |
| 961 | 8.8 | 0.0 | 2.1 | 5.4 | mttA/Hcf106 family, putative |
| 899 | 8.8 | 0.0 | 0.0 | 0.0 | riboflavin biosynthesis protein RibD |
| 878 | 8.8 | 0.0 | 0.0 | 0.0 | ABC transporter permease component |
| 836 | 8.8 | 0.0 | 0.0 | 0.0 | ribosomal protein L23 |
| 743 | 8.8 | 8.7 | 0.0 | 0.0 | conserved hypothetical protein |
| 707 | 8.8 | 0.0 | 0.0 | 0.0 | conserved hypothetical protein |
| 674 | 8.8 | 5.8 | 0.0 | 1.1 | ABC transporter, quaternary amine uptake transporter (QAT) family, substrate-binding protein, putative |
| 595 | 8.8 | 0.0 | 0.0 | 0.0 | conserved hypothetical protein |
| 589 | 8.8 | 0.0 | 0.4 | 0.0 | 3-oxoacyl-[acyl-carrier-protein] reductase |
| 472 | 8.8 | 0.0 | 0.0 | 0.0 | modification methylase |
| 377 | 8.8 | 0.0 | 0.0 | 0.0 | conserved hypothetical protein |
| 1337 | 8.8 | 5.8 | 0.0 | 0.0 | type II Secretion PilT |
| 1316 | 8.8 | 0.0 | 0.0 | 0.0 | glutaredoxin 3 |
| 1133 | 8.8 | 0.0 | 0.0 | 0.0 | glutathione-dependent formaldehyde-activating, GFA, putative |
| 1126 | 8.8 | 0.0 | 0.0 | 0.0 | sulfide dehydrogenase |
| 920 | 0.0 | 8.7 | 0.0 | 2.7 | 6-O-methylguanine DNA methyltransferase |
| 90 | 2.9 | 8.7 | 1.5 | 0.0 | translation initiation factor IF-1 |
| 42 | 0.0 | 8.7 | 0.0 | 0.0 | 3-oxoacyl-[acyl-carrier-protein] reductase, putative |
| 349 | 0.0 | 8.7 | 0.9 | 0.0 | acid tolerance regulatory protein actr |
| 30 | 0.0 | 8.7 | 0.0 | 0.0 | UDP-glucose 4-epimerase |
| 293 | 5.9 | 8.7 | 0.0 | 0.0 | mannitol transporter |
| 205 | 0.0 | 8.7 | 0.0 | 0.0 | putative porin |
| 1227 | 0.0 | 8.7 | 0.0 | 0.0 | conserved hypothetical protein |
| 1214 | 7.4 | 8.7 | 1.7 | 8.6 | substrate-binding region of ABC-type glycine betaine transport system |
| 1074 | 4.4 | 8.7 | 2.1 | 0.0 | serine--glyoxylate aminotransferase |
| 821 | 7.9 | 2.8 | 4.3 | 0.9 | translation elongation factor Tu |
| 687 | 7.6 | 7.8 | 4.6 | 2.7 | taurine transport system periplasmic protein |
| 1225 | 4.4 | 7.6 | 1.3 | 3.9 | ammonium transporter |
| 420 | 5.3 | 7.2 | 2.8 | 0.9 | ATP synthase subunit C, putative |
| 1129 | 6.6 | 7.1 | 7.0 | 3.6 | non-specific DNA-binding protein HBsu |
| 737 | 7.1 | 1.7 | 0.0 | 0.0 | trap dicarboxylate transporter- dctp subunit |
| 1269 | 2.2 | 2.9 | 0.4 | 7.0 | ABC proline/glycine betaine transporter, periplasmic substrate-binding protein |
| 855 | 6.6 | 2.5 | 0.0 | 0.0 | ribosomal protein S13p/S18e |
| 480 | 6.6 | 0.0 | 0.0 | 0.5 | cell division protein FtsZ |
| 96 | 0.0 | 0.0 | 6.2 | 0.0 | ribosomal protein L34 |
| 815 | 0.0 | 0.0 | 6.2 | 0.0 | prepilin-type N-terminal cleavage/methylation domain protein |
| 637 | 0.0 | 0.0 | 6.2 | 0.9 | molybdenum cofactor biosynthesis protein C |