

GHGT-11

## Metagenomics in CO<sub>2</sub> monitoring

Othilde Elise Håvelsrud<sup>a,b</sup>, Thomas H.A. Haverkamp<sup>c</sup>, Tom Kristensen<sup>b,d</sup>, Kjetill S. Jakobsen<sup>c,d</sup> and Anne Gunn Rike<sup>a,\*</sup>

<sup>a</sup>Norwegian Geotechnical Institute, P.O. Box 3930 Ullevål Stadion, N-0806 Oslo, Norway

<sup>b</sup>Department of Molecular Biosciences, University of Oslo, P.O. Box 1041 Blindern, N-0316 Oslo, Norway

<sup>c</sup>Centre for Evolutionary and Ecological Synthesis, CEES, Dept of Biology, University of Oslo, P.O. Box 1066 Blindern, N-0316 Oslo, Norway

<sup>d</sup>Microbial Evolution Research Group, MERG, Department of Biology, University of Oslo, Blindernveien 31, P.O. Box 1066 Blindern, N-0316 Oslo, Norway

### Abstract

Leakage from CO<sub>2</sub> storage areas is likely to affect the microbial communities in the overlying sediments. We have conducted a baseline characterization of the microbial communities present in the surface sediments overlying the Johansen formation, a potential site for CO<sub>2</sub> storage, using metagenomics. We detected six abundant potentially CO<sub>2</sub> fixing strains (e.g. *Nitrosopumilus maritimus* SCM1) as well as key genes for CO<sub>2</sub> fixation pathways (e.g. the reductive tricarboxylic acid cycle and the Wood Ljungdahl pathway). Assuming this fraction of the community would increase in case of CO<sub>2</sub> leakage; this information could be used as part of a surveillance project.

© 2013 The Authors. Published by Elsevier Ltd.  
Selection and/or peer-review under responsibility of GHGT

CCS; reservoir monitoring; metagenomics; marker genes; autothrops

### 1. Introduction

After the industrial revolution there has been a major increase in the atmospheric concentration of CO<sub>2</sub> and other green house gases [1]. A promising way to mitigate this development is CO<sub>2</sub> capture and storage (CCS), storing of CO<sub>2</sub> deep under the earth's surface [1]. During the lifetime of a geological CO<sub>2</sub> storage site diffuse migration or leakage of CO<sub>2</sub> from the reservoir to the overburden and sub seafloor may occur. As CCS increase in importance, sensitive and varied methods for leakage monitoring also increase in significance. Further, it is important to characterize expected alternations of the environment in case of a potential leakage.

Prokaryotic community structure and metabolic activity is tightly linked to geochemical parameters in the environment (e.g. carbon sources and red-ox conditions). Changes in prokaryotic sediment communities might therefore be among the first detectable warnings if a CO<sub>2</sub> leakage should occur. A recent study of soil microbial communities using 454 amplicon-sequencing and GeoChip analyses demonstrated dramatic changes in community structure and metabolic potential (e.g. significantly increased abundance in genes involved in CO<sub>2</sub> fixation) related to elevated CO<sub>2</sub> concentrations [2]. Both

\* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .  
E-mail address: [author@institute.xxx](mailto:author@institute.xxx) .

quantitative and qualitative changes in the microbial community after CO<sub>2</sub> injection into subsurface saline aquifers in Ketzin, Germany, have also been observed by FISH and fingerprinting techniques [3].

Over the last decades several studies on microbial communities in marine sediments, especially related to hotspots like methane seeps (e.g. the Coal Oil Point seepage area [4] and the Håkon Mosby mud volcano [5, 6], hydrothermal vents (e.g. Rainbow, Lost City and Ashadze [7] and whale falls [8], have been conducted. Still, little is known of communities associated with increased CO<sub>2</sub> levels. It is however known that several prokaryotes have the ability to assimilate CO<sub>2</sub> (inorganic carbon) into organic carbon. Currently there are six known pathways of CO<sub>2</sub> fixation [9]. Three of these are anaerobic or oxygen sensitive: the Wood Ljungdahl (WL) pathway, the reductive tricarboxylic acid (rTCA) cycle and the dicarboxylate/ 4-hydroxybutyrate (DC/ 4-HB) pathway. The other three are aerobic: the Calvin-Benson-Bassham (CBB) cycle, the 3-hydroxypropionate (3-HP) bicycle and the 3-hydroxypropionate/ 4-hydroxybutyrate (3-HP/4-HB) cycle. Baseline characterization of prokaryotic communities in sediments overlaying potential storage sites, as well as knowledge of their change in case of leakage is therefore of major importance.

It is assumed that less than 1% of existing prokaryotes has been cultivated in pure culture. Culture independent methods are therefore preferable. In this study we have used metagenomics to conduct a baseline characterization of the microbial communities present in the surface sediments overlaying the Johansen formation, a potential site for CO<sub>2</sub> storage in the North Sea. The goal of the study was to identify organisms and genes known to be involved in CO<sub>2</sub> fixation.

## 2. Methods

A flowchart of the methods used in this study can be seen in Fig. 1. The sediment samples were taken from the northern part of the North Sea from the area overlaying the Troll oil and gas field and a potential CO<sub>2</sub> storage site- the Johansen formation. The seafloor in this area is characterized by a high density of pockmarks and all our samples, except 1T, were taken from the bottom of pockmarks. The area has previously been investigated in relation to prokaryotic hydrocarbon degradation [10]. Details on the sampling locations are listed in Table 1.

The samples were collected using a combination of a 0.5 m ROV-operated shallow core device and a ROV manipulator. The ends of the core liners were sealed and kept at 4-10°C during transport.

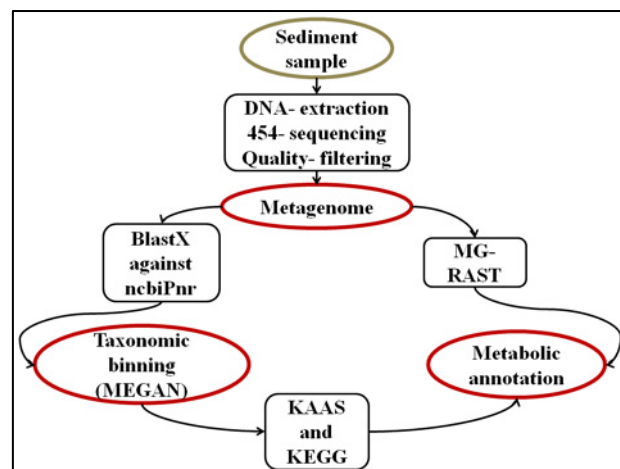


Fig. 1 Flowchart of the metagenomic analysis

Table 1 Sampling site locations

Sample	Latitude - longitude (N – E)	Water depth (m)	Sediment depth (cm bsf)	Sediment type
1T	60.631117 - 3.787293	305	5-20	Silty clay
2T	60.63132 - 3.789782	315	5-20	Silty clay
4T	60.631441 - 3.790041	315	5-20	Silty clay
6T	60.630721 - 3.78115	311	5-20	Silty clay
7T	60.629635 - 3.782211	311	5-15	Silty clay

### 2.1. DNA extraction and 454 sequencing

Total genomic DNA was extracted with a FastDNA®SPIN for Soil Kit (MP Biomedicals) and cleaned using Wizard DNA Clean-Up (Promega) according to the manufacturer's instructions. The DNA quality was assessed by agarose gel electrophoresis and by optical density using a NanoDrop instrument (NanoDrop Products, Thermo Scientific).

4-20 µg DNA was used for sequencing. Sample preparation and sequencing of the extracted DNA were performed at the High Throughput Sequencing Centre at CEES, University of Oslo (<http://www.sequencing.uio.no/>) according to standard GS FLX Titanium protocols.

### 2.2. Quality filtering

The complete datasets were analyzed with Prinseq to determine the sequence quality scores [11]. For each sample we performed quality filtering to remove low quality reads (reads containing  $\geq 10$  ambiguous bp, or homopolymers  $\geq 10$  bp) using mothur [12]. The remaining reads were checked for exact duplicates using an UiO in-house script. Artificial replicates were removed using cdhit-454 with standard settings except minimal identity was set to 98% [13].

### 2.3. Taxonomic annotation and rarefaction analysis

The metagenomic reads were taxonomically classified by BlastX against the NCBI non-redundant Protein Database (ncbiP-nr) [14]. The computation was performed at the freely available Bioportal computer service (<http://www.bioportal.uio.no/>) [14]. Maximum expectation-value was set to  $10^{-3}$ , maximum 25 alignments were reported per hit.

The BlastX output files were analyzed according to the NCBI-taxonomy in MEGAN, version 4 [15, 16] using default LCA-parameters (Min Score: 35, Top Percent: 10.0 and Min Support: 5). All taxa were enabled. Rarefaction analysis was performed for *Archaea* and *Bacteria* at the most detailed level of the NCBI taxonomy.

### 2.4. Metabolic annotation

Reads assigned to prokaryotic taxonomic groups in MEGAN were extracted and assigned to KEGG Orthology (KO)-numbers on KAAS (KEGG Automatic Annotation Server) [17] (<http://www.genome.ad.jp/tools/kaas/>) using default parameters for SBH (single-directional best hit). Forty reference genomes were manually selected based on the most abundant prokaryotic strains identified in the MEGAN-analysis.

KO numbers were assigned to the following taxonomic groups:

- Total *Archaea*
- *Nitrosopumilus maritimus* SCM1 (The most abundant archaeal strain.)
- Total *Bacteria*
- *Proteobacteria* (The most abundant bacterial phylum)
- "unclassified and environmental bacteria" (Made up of reads not assigned further than to *Bacteria* in addition to environmental samples (*Bacteria*) and unclassified bacteria). This group includes several BAC (bacterial artificial chromosomes) and fosmid based genomic sequences as well as some candidate divisions.
- "other bacterial phyla" (All bacterial nodes at the "phylum level" except *Proteobacteria*, unclassified bacteria and environmental samples).

The metagenomic reads were also assigned to SEED subsystems on the MG-RAST server (version 2.0) [18] (<http://metagenomics.anl.gov/v2/>). Maximum expectation- value was set to  $10^{-5}$ , minimum alignment length was set to 100 bp.

### 2.5. Effective Genome Size (EGS)

The effective genome size (EGS) for each metagenome was estimated according to the method developed by Raes et al. [19], using the constants  $a = 18.26$ ,  $b = 3650$  and  $c = 0.733$ . A protein reference database containing the 35 single copy COGs in question were downloaded from STRING (9.0) (<http://string-db.org/>) [19]. BlastX was conducted at the the freely available Biportal computer service (<http://www.biportal.uio.no/>) [14].

Sampling probability of a random universal single copy gene (1000 bases) and expected number of reads detected was calculated according to Beszteri et al. [20].

### 3. Results and Discussion

Five sediment metagenomes were sequenced to get an overview of the organisms and genes involved in CO<sub>2</sub> fixation. After replicate removal and quality filtering the metagenomes consisted of 607 557 to 1 227 131 reads, with an average read length of 336 to 361 (Table 2).

Unless other ways specified, all percentages in the following text are given as percent of total reads in each of the filtered metagenomes.

Table 2 Overview of the five metagenomes after replicate removal and quality filtering

Metagenome	Reads	Mean length (bp)	Mean GC (%)	EGS (Mbp)
1T	850 039	349	53.9	5.1
2T	663 131	361	49.9	4.7
4T	1 227 131	346	50.6	5.0
6T	607 557	343	49.3	4.6
7T	898 796	336	49.8	5.0

#### 3.1. Taxonomic annotation

Most of the reads were assigned to *Bacteria* (from 41 to 50%). The highest representation of *Archaea*, *Eukaryota* and Viruses were 3.5%, 1.75% and 0.17% respectively. Up to 43% of the reads had no hits against the ncbiP-nr.

Approximately 1000 prokaryotic (archaeal and bacterial) taxa (strains) were detected in each metagenome at the most resolved level in MEGAN. Rarefaction curves supported that the most abundant strains were accounted for, although our metagenomes did not capture the total prokaryotic richness in the sediments (Fig. 2).

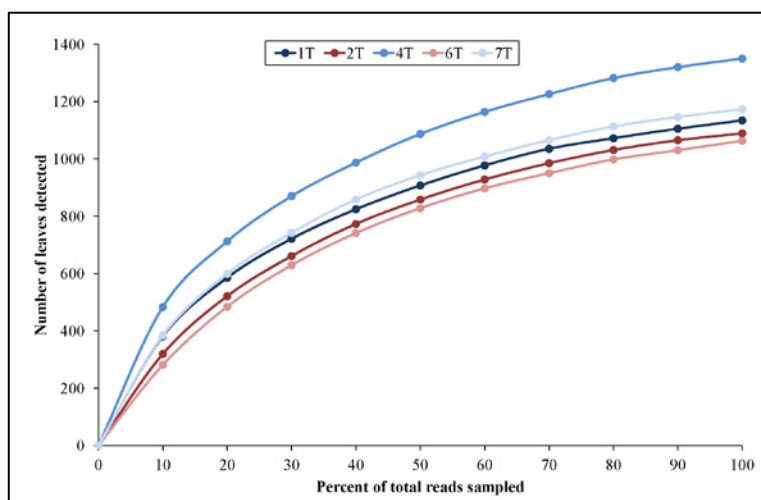


Fig. 2 Rarefaction curves for archaeal and bacterial strains

We focused our attention on the most abundant bacterial and archaeal strains in the sediments (more than 0.01% of all reads, when all metagenomes were put together) (Fig. 3). Among the 27 most abundant taxa identified in the MEGAN analysis, six are known to contribute to carbon fixation (Fig. 3). *Candidatus Nitrospira defluvii* uses the rTCA pathway, while the deltaproteobacterial taxa (*D. autotrophicum* HRM2, *D. alkenivorans* AK-01, and uncultured *Desulfobacterium*) and *Candidatus Kuenenia stuttgartiensis* utilizes the WL- pathway [9, 21-25]. Finally the archaeon *N. maritimus* likely makes use of a variant of the 3HP/4HB- pathway [26]. The abundance of these taxa, as well as other microbial groups able to use CO<sub>2</sub> as a carbon source, may increase as a response to elevated CO<sub>2</sub> concentrations.

### 3.2. Metabolic annotation

The metagenomic reads were annotated to metabolic functions in order to further assess the communities' potential for CO<sub>2</sub> fixation (Fig. 1). Most attention was given to the CO<sub>2</sub> fixation pathways assumed to be utilized by the most abundant taxa (Fig. 3).

#### 3.2.1. The Wood Ljungdahl pathway

Mapping reads to KEGG maps showed that the complete WL pathway was identified among the extracted proteobacterial reads (A.1.), supporting the taxonomic finding of the three abundant deltaproteobacterial taxa (*D. autotrophicum* HRM2, *D. alkenivorans* AK-01, and uncultured *Desulfobacterium*) assumed to use this pathway (Fig. 3).

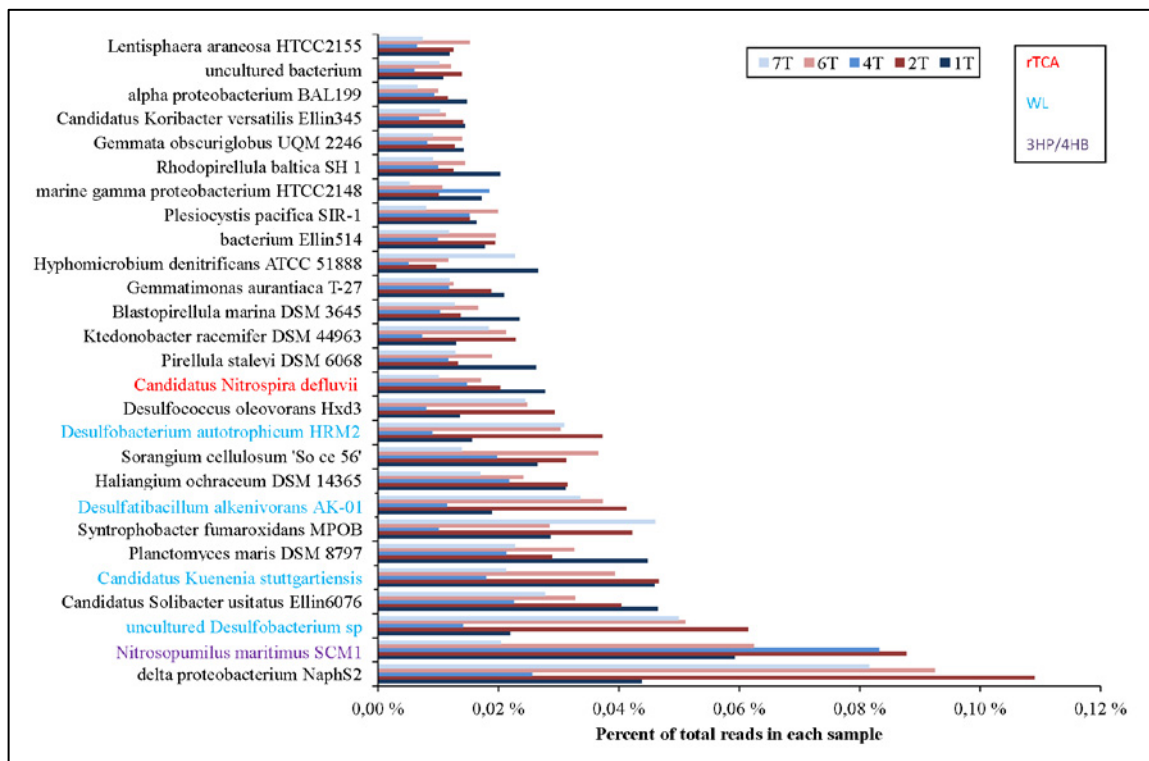


Fig. 3 Most abundant strains. Taxa potentially involved in CO<sub>2</sub> fixation and their pathways are indicated (rTCA: the reductive tricarboxylic acid cycle, WL: the Wood Ljungdahl pathway, 3HP/4HB: 3-hydroxypropionate/ 4-hydroxybutyrate cycle).

The complete WL pathway was also identified among the reads extracted from “other bacterial phyla” (including the abundant potential CO<sub>2</sub> fixing *Candidatus* Kuenina stuttgartiensis) and “unclassified and environmental bacteria”. In fact most reads assigned to the key enzyme, carbon monoxide dehydrogenase/ acetyl-CoA synthase (EC: 1.2.7.4, 1.2.99.2 and 2.3.1.169), were extracted from “unclassified and environmental bacteria” suggesting that this pathway is important among members of this group.

Reads assigned to euryarchaeotal methanogenic orders like *Methanosacinales* and *Methanomicrobiales*, known to use the WL-pathway, were identified in the MEGAN analysis. Still, four enzymes of the WL pathway remained undetected among the total archaeal reads using KEGG-maps. The absence of acetate kinase (EC 2.7.2.1), phosphate acetyltransferase (EC 2.3.1.8) and formate dehydrogenase (EC 1.2.1.43) is probably related to major differences in enzymes involved in this pathway for archaea and bacteria. The KEGG-map is based on the bacterial pathway, while intermediates like formate and methyltetrahydrofolate (and related enzymes) are replaced by formyl-methanofuran and tetrahydropterins in archaea [9]. Only the key enzyme, CO dehydrogenase/ acetyl-CoA synthase (EC: 1.2.7.4, 1.2.99.2 and 2.3.1.169), is believed to share a common origin in both archaea and bacteria [9, 27]. This enzyme was however also absent among the archaeal reads. The beta subunit of this key enzyme was not detected at all in our metagenomes. This could be due to the high taxonomic richness of prokaryotes in the Troll sediments, which combined with Effective Genome Sizes (EGS) of 4.6- 5.1 leads to low coverage of most genomes represented in the metagenomes (Fig. 2, Table 2). If a random example gene of 1000 bp was present in one copy in all organisms, between 181 and 199 hits to this gene could be expected in each metagenome [19, 20]. Genes present only in a small sub population might however be missed by chance.

Reads assigned to the SEED subsystem associated with the WL pathway were also detected in the MGRAST analysis (A.2.).

### 3.2.2. The reductive tricarboxylic acid cycle

All enzymes needed for the rTCA pathway were identified after mapping the reads assigned to “other bacterial phyla”, which includes the abundant strain *Candidatus* Nitrospira defluvii, to KEGG maps (A.3.). This group also contained reads assigned to other taxa known to use the rTCA pathway (e. g. taxa within *Chlorobi* and *Aquificae*) [9].

The rTCA pathway was further complete for “unclassified and environmental bacteria”, indicating the presence of organisms capable of rTCA among these groups as well.

Although the proteobacterial taxa identified in the MEGAN analysis included taxa reported to have an operating rTCA pathway (like *Magnetococcus* sp. MC-1 and *Candidatus* Endoriftia Persephone) the citrate cleavage enzyme (EC 2.3.3.8) could not be identified among reads assigned to this phylum. The postulated use of an alternative enzyme with the same function in some proteobacterial taxa (like *Magnetococcus* sp. MC-1 and *Candidatus* Endoriftia Persephone) could be a contributing factor to the absence of this enzyme [9].

This pathway is considered to be strictly bacterial and the complete pathway was not identified among the archaeal reads [9].

### 3.2.3. The 3-hydroxypropionate/ 4-hydroxybutyrate cycle

This pathway is expected to be used by archaea only [9]. Taxa expected to use this pathway were detected in the MEGAN analysis (*Sulfolobales*, *Cenarchaeum symbiosum* and *N. maritimus*) [9]. Still, enzymes for only some of the pathway steps were detected among our archaeal reads (A.4.). In this case it is important to keep in mind that a complete 3HP/4HB pathway has been identified in *Metallophaera sedula* only. Further, although the abundant taxon *N. maritimus* likely uses the same reaction sequences as *M. sedula*, not all reactions are catalyzed by identical enzymes [26]. The identification of 4-hydroxybutyryl-CoA dehydratase/vinylacetyl-CoA-Delta-isomerase, a characteristic key gene of the

3HP/4HB cycle, among the reads assigned to *N. maritimus* does however support genetic potential for CO<sub>2</sub> fixation [9].

#### 3.2.4. Other CO<sub>2</sub> fixation pathways

Only low abundant taxa known to use the DC/ 4-HB pathway (e.g. *Desulfurococcales* and *Thermoproteales*) and the 3-HP pathway (*Chloroflexaceae*) were detected in the MEGAN analysis. Further, the complete pathways were not identified by KEGG mapping (A.5. and A.6.). These pathways are therefore likely not of major importance in the sediments at the present time.

None of the most abundant taxa in the sediments are known to use the CBB cycle and not all enzymes in the KEGG pathway could be identified (A.7.). Still, annotation to level III SEED subsystems on MG-RAST indicated that the CBB cycle and its related CO<sub>2</sub> uptake system and carboxysomes (organelle-like proteinaceous polyhedral micro-compartments thought to facilitate carbon fixation via the CBB-cycle) was important in CO<sub>2</sub> fixation in the Troll sediments (A.2.). This inconsistency could be due to differences in the respective databases, as well as differences in the classification of enzymes sorting under each pathway/subsystem (e.g. the carboxysome genes are not included in the KEGG-map).

Mapping of reads to KEGG pathway maps indicates that WL and rTCA are the most used pathways for CO<sub>2</sub> fixation in the sediments at the present time. Further, the high abundance of reads assigned to *N. maritimus* in combination with the detection of 4-hydroxybutyryl-CoA dehydratase/vinylacetyl-CoA-Delta-isomerase indicates that a version of the 3HP/4HB cycle may also be operational. In addition SEED subsystems related to CBB were indicated to be important in the MG-RAST analysis. It is therefore likely to assume an increased abundance of organisms using these pathways, as well as genes needed for the pathways, should a leakage of CO<sub>2</sub> from the potential storage reservoir occur.

Sediments with their stratified and often steep chemical gradients constitute a complex and diverse habitat for the prokaryotes. Although we detected differences in the relative abundance of enzymes and autotrophic taxa between the samples we did not identify any patterns suggesting that certain pathways were more or less abundant in one sample compared to the others.

## 4. Summary and outlook

In sediment surveys with the aim to detect CO<sub>2</sub> levels beyond baseline one may search for changes in community profiles toward increased representation of autotrophic prokaryotes (e.g. *N. maritimus* SCM1, uncultured *Desulfobacterium* sp, *Candidatus* Kuenenia stuttgartiensis, *Desulfatibacillum alkanivorans* AK-01, *Desulfobacterium autotrophicum* HRM2, *Candidatus Nitrospira defluvii*). Both key genes in CO<sub>2</sub> fixation pathways (e.g. citric cleavage enzyme (rTCA), carbon monoxide dehydrogenase/acetyl-CoA synthase (WL)) and marker genes for the strains performing them (e.g. 16S rDNA) could be monitored.

Our study shows that metagenomic analyses represent a novel approach for monitoring of CO<sub>2</sub> induced changes in marine sediments. The high throughput next generation sequencing platforms, like the Roche/454 technology [28], renders an unprecedented sequencing coverage of environmental metagenomes possible. Metagenomics may therefore provide an early warning for leakage incidents. Metagenomics may further enable us to identify good marker genes for use in e.g. Q-PCR assays. Our work was carried out on marine sediments, but the same methods could also be applied in other environments, like surveillance of land-based CCS facilities.



## Acknowledgements

The metagenome sequencing and initial work was granted by VISTA/Statoil. OEH and the analytical costs were financed by project 6151 to AGR and THAH was financed by project 6503 to KSJ. The analyses directed towards CO<sub>2</sub> fixation was founded by SUCCESS Centre for CO<sub>2</sub> storage (grant 193825/S60 Research Council of Norway). The project was also supported by Norwegian Geotechnical Institutes education fund. The core samples and geochemical data were collected by the Norwegian Geotechnical Institute, in the Petrogen project (NFR 163467/S30, granted by the Research Council of Norway), and kindly provided to our metagenome project.

## References

- [1] R. Shukla, P. Ranjith, A. Haque, X. Choi, A review of studies on CO<sub>2</sub> sequestration and caprock integrity, *Fuel*, 2010; 89 2651-2664.
- [2] Z.L. He, M.Y. Xu, Y. Deng, S.H. Kang, L. Kellogg, L.Y. Wu, J.D. Van Nostrand, S.E. Hobbie, P.B. Reich, J.Z. Zhou, Metagenomic analysis reveals a marked divergence in the structure of belowground microbial communities at elevated CO<sub>2</sub>, *Ecology Letters*, 2010; 13 564-575.
- [3] D. Morozova, M. Zettlitzer, D. Let, H. Würdemann, the CO<sub>2</sub>Sink group, Monitoring of the microbial community composition in deep subsurface saline aquifers during CO<sub>2</sub> storage in Ketzin, Germany, In: J. Gale, C. Hendriks, W. Turkenberg (Eds.) *10th International Conference on Greenhouse Gas Control Technologies*, 2011, pp. 4362-4370.
- [4] O.E. Håvelsrud, T. Haverkamp, T. Kristensen, K. Jakobsen, A.G. Rike, A metagenomic study of methanotrophic microorganisms in Coal Oil Point seep sediments, *BMC Microbiology*, 2011; 11 221.
- [5] T. Lösekann, K. Knittel, T. Nadalig, B. Fuchs, H. Niemann, A. Boetius, R. Amann, Diversity and abundance of aerobic and anaerobic methane oxidizers at the Haakon Mosby Mud Volcano, Barents Sea, *Applied and environmental microbiology*, 2007; 73 3348-3362.
- [6] H. Niemann, T. Lösekann, D. de Beer, M. Elvert, T. Nadalig, K. Knittel, R. Amann, E.J. Sauter, M. Schlüter, M. Klages, J.P. Foucher, A. Boetius, Novel microbial communities of the Haakon Mosby mud volcano and their role as a methane sink, *Nature*, 2006; 443 854-858.
- [7] E.G. Roussel, C. Kohn, J.L. Charlou, J.P. Donval, Y. Fouquet, J. Querellou, D. Prieur, M.A.C. Bonavita, Comparison of microbial communities associated with three Atlantic ultramafic hydrothermal systems, *FEMS Microbiology Ecology*, 2011; 77 647-665.
- [8] S.K. Goffredi, V.J. Orphan, Bacterial community shifts in taxa and diversity in response to localized organic loading in the deep sea, *Environmental Microbiology*, 2010; 12 344-363.
- [9] M. Hügler, S.M. Sievert, Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean, In: C.A. Carlson, S.J. Giovannoni (Eds.) *Annual Review of Marine Science*, 2011, pp. 261-289.
- [10] O.E. Håvelsrud, T. Haverkamp, T. Kristensen, K.S. Jakobsen, A.G. Rike, Metagenomic and geochemical characterization of pockmarked sediments overlaying the Troll petroleum reservoir in the North Sea, *BMC Microbiology*, 2012; 12.
- [11] R. Schmieder, R. Edwards, Quality control and preprocessing of metagenomic datasets, *Bioinformatics*, 2011; 27 863-864.
- [12] P.D. Schloss, S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A. Lesniewski, B.B. Oakley, D.H. Parks, C.J. Robinson, J.W. Sahl, B. Stres, G.G. Thallinger, D.J. Van Horn, C.F. Weber, Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities, *Applied and Environmental Microbiology*, 2009; 75 7537-7541.
- [13] B.F. Niu, L.M. Fu, S.L. Sun, W.Z. Li, Artificial and natural duplicates in pyrosequencing reads of metagenomic data, *BMC Bioinformatics*, 2010; 11.
- [14] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *Journal of Molecular Biology*, 1990; 215 403-410.

- [15] D.H. Huson, A.F. Auch, J. Qi, S.C. Schuster, MEGAN analysis of metagenomic data, *Genome Research*, 2007; 17 377-386.
- [16] D.H. Huson, S. Mitra, H.J. Ruscheweyh, N. Weber, S.C. Schuster, Integrative analysis of environmental sequences using MEGAN4, *Genome Research*, 2011.
- [17] Y. Moriya, M. Itoh, S. Okuda, A.C. Yoshizawa, M. Kanehisa, KAAS: an automatic genome annotation and pathway reconstruction server, *Nucleic Acids Research*, 2007; 35 W182-W185.
- [18] F. Meyer, D. Paarmann, M. D'Souza, R. Olson, E.M. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens, A. Wilke, J. Wilkening, R.A. Edwards, The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes, *BMC Bioinformatics*, 2008; 9 386.
- [19] J. Raes, J.O. Korb, M.J. Lercher, C. von Mering, P. Bork, Prediction of effective genome size in metagenomic samples, *Genome Biology*, 2007; 8 R10.
- [20] B. Beszteri, B. Temperton, S. Frickenhaus, S.J. Giovannoni, Average genome size: a potential source of bias in comparative metagenomics, *ISME Journal*, 2010; 4 1075-1077.
- [21] S. Lückner, M. Wagner, F. Maixner, E. Pelletier, H. Koch, B. Vacherie, T. Rattei, J.S.S. Damsté, E. Spieck, D. Le Paslier, H. Daims, A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria, *Proceedings of the National Academy of Sciences of the United States of America*, 2010; 107 13479-13484.
- [22] J. Amann, D. Lange, M. Schüler, R. Rabus, Substrate-Dependent Regulation of Carbon Catabolism in Marine Sulfate-Reducing *Desulfobacterium autotrophicum* HRM2, *Journal of Molecular Microbiology and Biotechnology*, 2010; 18 74-84.
- [23] F. Bak, F. Widdel, Anaerobic degradation of indolic compounds by sulfate-reducing enrichment cultures, and description of *Desulfobacterium indolicum* gen. nov., sp. nov., *Archives of Microbiology*, 1986; 146 170-176.
- [24] A.V. Callaghan, B.E. Morris, I.A. Pereira, M.J. McInerney, R.N. Austin, J.T. Groves, J.J. Kukor, J.M. Suflita, L.Y. Young, G.J. Zylstra, B. Wawrik, The genome sequence of *Desulfatibacillum alkenivorans* AK-01: a blueprint for anaerobic alkane oxidation, *Environ Microbiol*, 2011.
- [25] M. Strous, E. Pelletier, S. Mangenot, T. Rattei, A. Lehner, M.W. Taylor, M. Horn, H. Daims, D. Bartol-Mavel, P. Wincker, V. Barbe, N. Fonknechten, D. Vallenet, B. Segurens, C. Schenowitz-Truong, C. Médigue, A. Collingro, B. Snel, B.E. Dutilh, H.J.M. Op den Camp, C. van der Drift, I. Cirpus, K.T. van de Pas-Schoonen, H.R. Harhangi, L. van Niftrik, M. Schmid, J. Keltjens, J. van de Vossenberg, B. Kartal, H. Meier, D. Frishman, M.A. Huynen, H.W. Mewes, J. Weissenbach, M.S.M. Jetten, M. Wagner, D. Le Paslier, Deciphering the evolution and metabolism of an anammox bacterium from a community genome, *Nature*, 2006; 440 790-794.
- [26] C.B. Walker, J.R. de la Torre, M.G. Klotz, H. Urakawa, N. Pinel, D.J. Arp, C. Brochier-Armanet, P.S.G. Chain, P.P. Chan, A. Gollabgir, J. Hemp, M. Hügler, E.A. Karr, M. Könneke, M. Shin, T.J. Lawton, T. Lowe, W. Martens-Habbena, L.A. Sayavedra-Soto, D. Lang, S.M. Sievert, A.C. Rosenzweig, G. Manning, D.A. Stahl, *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea, *Proceedings of the National Academy of Sciences of the United States of America*, 2010; 107 8818-8823.
- [27] I.A. Berg, D. Kockelkorn, W.H. Ramos-Vera, R.F. Say, J. Zarzycki, M. Hügler, B.E. Alber, G. Fuchs, Autotrophic carbon fixation in archaea, *Nature Reviews Microbiology*, 2010; 8 447-460.
- [28] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z.T. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L.I. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P.G. Yu, R.F. Begley, J.M. Rothberg, Genome sequencing in microfabricated high-density picolitre reactors, *Nature*, 2005; 437 376-380.

## Appendix A. Tables

## A.1. Taxonomic distribution of reads assigned to functions in the Wood-Ljungdahl pathway

Enzyme	Total Archaea	N. Maritimus	Total Bacteria	Unclassified Bacteria	Proteobacteria	Other bacterial phyla
formate dehydrogenase alpha subunit (EC:1.2.1.43)	0	0	0	0	0	0
formate dehydrogenase beta subunit (EC:1.2.1.43)	0	0	150	97	34	19
putative pyruvate-flavodoxin oxidoreductase (EC:1.2.7.-)	0	0	921	783	98	40
pyruvate ferredoxin oxidoreductase, alpha subunit (EC:1.2.7.1)	51	0	260	123	59	78
pyruvate ferredoxin oxidoreductase, beta subunit (EC:1.2.7.1)	71	0	174	99	48	27
pyruvate ferredoxin oxidoreductase, delta subunit (EC:1.2.7.1)	15	0	49	16	23	10
pyruvate ferredoxin oxidoreductase, gamma subunit (EC:1.2.7.1)	58	0	96	39	38	19
carbon monoxide dehydrogenase / acetyl-CoA synthase subunit alpha (EC:1.2.7.4 1.2.99.2.2.3.1.169)	64	0	446	315	95	36
carbon monoxide dehydrogenase / acetyl-CoA synthase subunit beta (EC:1.2.7.4 1.2.99.2.2.3.1.169)	0	0	0	0	0	0
5-methyltetrahydrofolate corrinoid/iron sulfur protein methyltransferase (EC:2.3.1.169)	0	0	39	26	3	10
CO-methylating acetyl-CoA synthase (EC:2.3.1.169)	0	0	0	0	0	0
methylentetrahydrofolate reductase (NADPH) (EC:1.5.1.20)	9	0	632	272	287	73
methylentetrahydrofolate dehydrogenase (NADP+) / methenyltetrahydrofolate cyclohydrolase (EC:1.5.1.5.3.5.4.9)	20	10	424	276	106	42
phosphate acetyltransferase (EC:2.3.1.8)	0	0	84	63	12	9
phosphate acetyltransferase (EC:2.3.1.8)	0	0	177	125	27	25
propanediol utilization protein (EC:2.3.1.8)	0	0	0	0	0	0
acetate kinase (EC:2.7.2.1)	0	0	244	160	56	28
formate--tetrahydrofolate ligase (EC:6.3.4.3)	2	0	391	271	83	37

*A.2. Reads assigned to SEED subsystems involved in CO<sub>2</sub> fixation on MG-RAST (numbers are given as percent of total reads in each metagenome).*

Subsystem Name	1T	2T	4T	6T	7T
CO <sub>2</sub> uptake, carboxysome	0.0486	0.0437	0.0425	0.0304	0.0375
Carboxysome	0.0001	0.0003	0.0002	0.0000	0.0002
Calvin-Benson cycle	0.0173	0.0208	0.0218	0.0199	0.0218
Photorespiration (oxidative C <sub>2</sub> cycle)	0.0053	0.0051	0.0044	0.0044	0.0051
Wood-Ljungdahl pathway of CO <sub>2</sub> fixation	0.0002	0.0026	0.0006	0.0016	0.0011

*A.3. Taxonomic distribution of reads assigned to functions in the Reductive Tricarboxylic Acid Cycle*

Enzyme	Total Archaea	N. maritimus	Total Bacteria	Unclassified Bacteria	Proteobacteria	Other Bacterial phyla
malate dehydrogenase (EC:1.1.1.37)	36	8	409	250	104	55
isocitrate dehydrogenase (EC:1.1.1.42)	24	11	572	452	107	13
pyruvate ferredoxin oxidoreductase, alpha subunit (EC:1.2.7.1)	51	0	260	123	59	78
pyruvate ferredoxin oxidoreductase, beta subunit (EC:1.2.7.1)	71	0	174	99	48	27
pyruvate ferredoxin oxidoreductase, delta subunit (EC:1.2.7.1)	15	0	49	16	23	10
pyruvate ferredoxin oxidoreductase, gamma subunit (EC:1.2.7.1)	58	0	96	39	38	19
2-oxoglutarate ferredoxin oxidoreductase subunit alpha (EC:1.2.7.3)	118	35	1312	749	404	159
2-oxoglutarate ferredoxin oxidoreductase subunit beta (EC:1.2.7.3)	65	2	808	480	242	86
2-oxoglutarate ferredoxin oxidoreductase subunit delta (EC:1.2.7.3)	2	0	117	47	56	14
2-oxoglutarate ferredoxin oxidoreductase subunit gamma (EC:1.2.7.3)	9	0	168	68	74	26
succinate dehydrogenase flavoprotein subunit (EC:1.3.99.1)	70	10	959	531	303	125
succinate dehydrogenase iron-sulfur protein (EC:1.3.99.1)	40	13	454	258	144	52
succinate dehydrogenase cytochrome b-556 subunit (EC:1.3.99.1)	10	4	118	31	60	27
succinate dehydrogenase hydrophobic membrane anchor protein (EC:1.3.99.1)	0	0	25	5	19	1
fumarate reductase flavoprotein subunit (EC:1.3.99.1)	2	0	122	34	81	7

Enzyme	Total Archaea	N. maritimus	Total Bacteria	Unclassified Bacteria	Proteobacteria	Other bacterial phyla
fumarate reductase iron-sulfur protein (EC:1.3.99.1)	0	0	32	9	23	0
fumarate reductase subunit C (EC:1.3.99.1)	0	0	22	4	18	0
fumarate reductase subunit D (EC:1.3.99.1)	0	0	0	0	0	0
pyruvate, water dikinase (EC:2.7.9.2)	60	0	742	273	378	91
phosphoenolpyruvate carboxylase (EC:4.1.1.31)	0	0	247	68	143	36
ATP citrate (pro-S)-lyase (EC:2.3.3.8)	1	0	18	4	0	14
fumarate hydratase, class I (EC:4.2.1.2)	6	0	252	139	81	32
fumarate hydratase subunit alpha (EC:4.2.1.2)	16	0	101	72	14	15
fumarate hydratase subunit beta (EC:4.2.1.2)	1	0	57	48	6	3
fumarate hydratase, class II (EC:4.2.1.2)	31	18	376	222	82	72
aconitate hydratase 1 (EC:4.2.1.3)	62	38	890	590	229	71
aconitate hydratase 2 (EC:4.2.1.3)	1	0	66	22	35	9
succinyl-CoA synthetase alpha subunit (EC:6.2.1.5)	39	14	422	215	142	65
succinyl-CoA synthetase beta subunit (EC:6.2.1.5)	55	28	559	258	216	85
putative pyruvate-flavodoxin oxidoreductase (EC:1.2.7.-)	0	0	921	783	98	40

#### A.4. Taxonomic distribution of reads assigned to functions in the 3-Hydroxypropionate/ 4-Hydroxybutyrate Cycle

Enzyme	Total Archaea	N. maritimus	Total Bacteria	Unclassified Bacteria	Proteobacteria	Other bacterial phyla
succinate semialdehyde reductase (NADPH) (EC:1.1.1.-)	0	0	0	0	0	0
3-hydroxypropionate dehydrogenase (NADP+) (EC:1.1.1.298)	0	0	0	0	0	0
3-hydroxyacyl-CoA dehydrogenase (EC:1.1.1.35)	0	0	465	191	215	59
3-hydroxyacyl-CoA dehydrogenase / enoyl-CoA hydratase / 3-hydroxybutyryl-CoA epimerase (EC:1.1.1.35 4.2.1.17 5.1.2.3)	1	0	416	100	286	30
3-hydroxyacyl-CoA dehydrogenase / enoyl-CoA hydratase / 3-hydroxybutyryl-CoA epimerase / enoyl-CoA isomerase	0	0	3	0	3	0

Enzyme	Total Archaea	N. maritimus	Total Bacteria	Unclassified Bacteria	Proteobacteria	Other bacterial phyla
(EC:1.1.35.4.2.1.17.5.1.2.3.5.3.3.8)						
succinyl-CoA reductase (NADPH) (EC:1.2.1.-)	0	0	0	0	0	0
malonyl-CoA reductase / 3-hydroxypropionate dehydrogenase (NADP+) (EC:1.2.1.75.1.1.1.298)	0	0	0	0	0	0
succinyl-coA reductase (EC:1.2.1.76)	0	0	0	0	0	0
acryloyl-coenzyme A reductase (EC:1.3.1.84)	0	0	0	0	0	0
acrylyl-CoA reductase (NADPH) / 3-hydroxypropionyl-CoA dehydratase / 3- hydroxypropionyl-CoA synthetase (EC:1.3.1.84.4.2.1.116.6.2.1.36)	0	0	6	6	0	0
acetyl-CoA C-acetyltransferase (EC:2.3.1.9)	236	3	1756	682	844	230
3-hydroxypropionyl-coenzyme A dehydratase (EC:4.2.1.116)	0	0	0	0	0	0
4-hydroxybutyryl-CoA dehydratase / vinylacetyl-CoA-Delta-isomerase (EC:4.2.1.120.5.3.3.3)	49	13	187	112	53	22
3-hydroxybutyryl-CoA dehydratase (EC:4.2.1.55)	28	13	388	193	154	41
3-hydroxybutyryl-CoA dehydratase / 3-hydroxyacyl-CoA dehydrogenase (EC:4.2.1.55.1.1.1.35)	0	0	0	0	0	0
methylmalonyl-CoA epimerase (EC:5.1.99.1)	7	2	145	69	60	16
methylmalonyl-CoA mutase (EC:5.4.99.2)	0	0	338	192	122	24
methylmalonyl-CoA mutase (EC:5.4.99.2)	0	0	481	415	38	28
methylmalonyl-CoA mutase, C-terminal domain (EC:5.4.99.2)	12	2	161	126	31	4
methylmalonyl-CoA mutase, N-terminal domain (EC:5.4.99.2)	48	22	495	266	171	58
4-hydroxybutyryl-CoA synthetase (4-hydroxybutyrate-CoA ligase, AMP-forming) (EC:6.2.1.-)	0	0	0	0	0	0
4-hydroxybutyryl-CoA synthetase (4-hydroxybutyrate-CoA ligase, AMP-forming) (EC:6.2.1.-)	0	0	0	0	0	0
3-hydroxypropionyl-coenzyme A synthetase (EC:6.2.1.36)	0	0	0	0	0	0
acetyl-CoA carboxylase biotin carboxyl carrier protein (EC:6.4.1.2)	0	0	116	62	48	6
acetyl-CoA carboxylase carboxyl transferase subunit alpha (EC:6.4.1.2)	0	0	350	171	137	42
acetyl-CoA carboxylase carboxyl transferase subunit beta (EC:6.4.1.2)	0	0	297	145	112	40
acetyl-CoA carboxylase, biotin carboxylase subunit (EC:6.4.1.2.6.3.4.14)	29	0	895	466	354	75
acetyl-CoA/propionyl-CoA carboxylase (EC:6.4.1.2.6.4.1.3)	0	0	0	0	0	0

Enzyme	Total Archaea	N. maritimus	Total Bacteria	Unclassified Bacteria	Proteobacteria	Other bacterial phyla
acetyl-CoA:propionyl-CoA carboxylase (EC:6.4.1.2 6.4.1.3)	0	0	0	0	0	0
biotin carboxyl carrier protein (EC:6.4.1.2 6.4.1.3)	0	0	0	0	0	0
propionyl-CoA carboxylase (EC:6.4.1.3)	0	0	8	5	2	1
propionyl-CoA carboxylase alpha chain (EC:6.4.1.3)	0	0	29	1	28	0
propionyl-CoA carboxylase beta chain (EC:6.4.1.3)	47	9	569	336	197	36

#### A.5. Taxonomic distribution of reads assigned to functions in the Dicarboxylate/4-Hydroxybuturate Cycle

Enzyme	Total Archaea	N. maritimus	Total Bacteria	Unclassified Bacteria	Proteobacteria	Other Bacterial phyla
malate dehydrogenase (EC:1.1.1.37)	36	8	409	250	104	55
pyruvate ferredoxin oxidoreductase, alpha subunit (EC:1.2.7.1)	51	0	260	123	59	78
pyruvate ferredoxin oxidoreductase, beta subunit (EC:1.2.7.1)	71	0	174	99	48	27
pyruvate ferredoxin oxidoreductase, delta subunit (EC:1.2.7.1)	15	0	49	16	23	10
pyruvate ferredoxin oxidoreductase, gamma subunit (EC:1.2.7.1)	58	0	96	39	38	19
succinate dehydrogenase flavoprotein subunit (EC:1.3.99.1)	70	10	959	531	303	125
succinate dehydrogenase iron-sulfur protein (EC:1.3.99.1)	40	13	454	258	144	52
succinate dehydrogenase cytochrome b-556 subunit (EC:1.3.99.1)	10	4	118	31	60	27
succinate dehydrogenase hydrophobic membrane anchor protein (EC:1.3.99.1)	0	0	25	5	19	1
fumarate reductase flavoprotein subunit (EC:1.3.99.1)	2	0	122	34	81	7
fumarate reductase iron-sulfur protein (EC:1.3.99.1)	0	0	32	9	23	0
fumarate reductase subunit C (EC:1.3.99.1)	0	0	22	4	18	0
fumarate reductase subunit D (EC:1.3.99.1)	0	0	0	0	0	0
acetyl-CoA C-acetyltransferase (EC:2.3.1.9)	236	3	1756	682	844	230
pyruvate, water dikinase (EC:2.7.9.2)	60	0	742	273	378	91

Enzyme	Total Archaea	N. maritimus	Total Bacteria	Unclassified Bacteria	Proteobacteria	Other Bacterial phyla
phosphoenolpyruvate carboxylase (EC:4.1.1.31)	0	0	247	68	143	36
fumarate hydratase, class I (EC:4.2.1.2)	6	0	252	139	81	32
fumarate hydratase subunit alpha (EC:4.2.1.2)	16	0	101	72	14	15
fumarate hydratase subunit beta (EC:4.2.1.2)	1	0	57	48	6	3
fumarate hydratase, class II (EC:4.2.1.2)	31	18	376	222	82	72
3-hydroxybutyryl-CoA dehydratase (EC:4.2.1.55)	28	13	388	193	154	41
3-hydroxyacyl-CoA dehydrogenase / enoyl-CoA hydratase / 3-hydroxybutyryl-CoA epimerase (EC:1.1.1.35 4.2.1.17 5.1.2.3)	1	0	416	100	286	30
3-hydroxyacyl-CoA dehydrogenase / enoyl-CoA hydratase / 3-hydroxybutyryl-CoA epimerase / enoyl-CoA isomerase (EC:1.1.1.35 4.2.1.17 5.1.2.3 5.3.3.8)	0	0	3	0	3	0
succinyl-CoA synthetase alpha subunit (EC:6.2.1.5)	39	14	422	215	142	65
succinyl-CoA synthetase beta subunit (EC:6.2.1.5)	55	28	559	258	216	85
putative pyruvate-flavodoxin oxidoreductase (EC:1.2.7.-)	0	0	921	783	98	40
3-hydroxyacyl-CoA dehydrogenase (EC:1.1.1.35)	0	0	465	191	215	59
succinate semialdehyde reductase (NADPH) (EC:1.1.1.-)	0	0	0	0	0	0
4-hydroxybutyryl-CoA synthetase (4-hydroxybutyrate-CoA ligase, AMP-forming) (EC:6.2.1.-)	0	0	0	0	0	0
4-hydroxybutyryl-CoA synthetase (4-hydroxybutyrate-CoA ligase, AMP-forming) (EC:6.2.1.-)	0	0	0	0	0	0
4-hydroxybutyryl-CoA dehydratase / vinylacetyl-CoA-Delta-isomerase (EC:4.2.1.120 5.3.3.3)	49	13	187	112	53	22
3-hydroxybutyryl-CoA dehydratase / 3-hydroxyacyl-CoA dehydrogenase (EC:4.2.1.55 1.1.1.35)	0	0	0	0	0	0
succinyl-coA reductase (EC:1.2.1.76)	0	0	0	0	0	0



## A.6. Taxonomic distribution of reads assigned to functions in the 3-Hydroxypropionate Bicycle

Enzyme	Total Archaea	N. maritimus	Total Bacteria	Unclassified Bacteria	Proteobacteria	Other bacterial phyla
3-hydroxypropionate dehydrogenase (NADP+) (EC:1.1.1.298)	0	0	0	0	0	0
succinyl-CoA reductase (NADPH) (EC:1.2.1.1-)	0	0	0	0	0	0
malonyl-CoA reductase / 3-hydroxypropionate dehydrogenase (NADP+) (EC:1.2.1.75 1.1.1.298)	0	0	0	0	0	0
acryloyl-coenzyme A reductase (EC:1.3.1.84)	0	0	0	0	0	0
acrylyl-CoA reductase (NADPH) / 3-hydroxypropionyl-CoA dehydratase / 3- hydroxypropionyl-CoA synthetase (EC:1.3.1.84 4.2.1.116 6.2.1.36)	0	0	6	6	0	0
fumarate reductase subunit C (EC:1.3.91.1)	0	0	22	4	18	0
fumarate reductase subunit D (EC:1.3.91.1)	0	0	0	0	0	0
succinate dehydrogenase cytochrome b-556 subunit (EC:1.3.91.1)	10	4	118	31	60	27
succinate dehydrogenase hydrophobic membrane anchor protein (EC:1.3.91.1)	0	0	25	5	19	1
fumarate reductase flavoprotein subunit (EC:1.3.99.1)	2	0	122	34	81	7
fumarate reductase iron-sulfur protein (EC:1.3.99.1)	0	0	32	9	23	0
succinate dehydrogenase flavoprotein subunit (EC:1.3.99.1)	70	10	959	531	303	125
succinate dehydrogenase iron-sulfur protein (EC:1.3.99.1)	40	13	454	258	144	52
succinyl-CoA:(S)- malate CoA transferase subunit A (EC:2.8.3.-)	1	0	12	12	0	0
succinyl-CoA:(S)- malate CoA transferase subunit B (EC:2.8.3.-)	1	0	11	8	2	1
malyl-CoA lyase (EC:4.1.3.24)	0	0	75	27	16	32
3-hydroxypropionyl-coenzyme A dehydratase (EC:4.2.1.116)	0	0	0	0	0	0
fumarate hydrate subunit alpha (EC:4.2.1.2)	16	0	101	72	14	15
fumarate hydrate subunit beta (EC:4.2.1.2)	1	0	57	48	6	3
fumarate hydrate, class I (EC:4.2.1.2)	6	0	252	139	81	32
fumarate hydrate, class II (EC:4.2.1.2)	31	18	376	222	82	72
methylmalonyl-CoA epimerase (EC:5.1.99.1)	7	2	145	69	60	16
methylmalonyl-CoA mutase (EC:5.4.99.2)	0	0	338	192	122	24

Enzyme	Total Archaea	N. maritimus	Total Bacteria	Unclassified Bacteria	Proteobacteria	Other bacterial phyla
methylmalonyl-CoA mutase (EC:5.4.99.2)	0	0	481	415	38	28
methylmalonyl-CoA mutase, C-terminal domain (EC:5.4.99.2)	12	2	161	126	31	4
methylmalonyl-CoA mutase, N-terminal domain (EC:5.4.99.2)	48	22	495	266	171	58
3-hydroxypropionyl-coenzyme A synthetase (EC:6.2.1.36)	0	0	0	0	0	0
acetyl-CoA carboxylase biotin carboxyl carrier protein (EC:6.4.1.2)	0	0	116	62	48	6
acetyl-CoA carboxylase carboxyl transferase subunit alpha (EC:6.4.1.2)	0	0	350	171	137	42
acetyl-CoA carboxylase carboxyl transferase subunit beta (EC:6.4.1.2)	0	0	297	145	112	40
acetyl-CoA carboxylase, biotin carboxylase subunit (EC:6.4.1.2 6.3.4.14)	29	0	895	466	354	75
acetyl-CoA/propionyl-CoA carboxylase (EC:6.4.1.2 6.4.1.3)	0	0	0	0	0	0
acetyl-CoA/propionyl-CoA carboxylase (EC:6.4.1.2 6.4.1.3)	0	0	0	0	0	0
biotin carboxyl carrier protein (EC:6.4.1.2 6.4.1.3)	0	0	0	0	0	0
propionyl-CoA carboxylase (EC:6.4.1.3)	0	0	8	5	2	1
propionyl-CoA carboxylase alpha chain (EC:6.4.1.3)	0	0	29	1	28	0
propionyl-CoA carboxylase beta chain (EC:6.4.1.3)	47	9	569	336	197	36
mesaconyl-CoA hydratase (Meh)	0	0	66	43	22	1
mesaconyl-CoA C1-C4 CoA transferase (Mct)	0	0	24	17	3	4
mesaconyl-C4 CoA hydratase (Meh)	0	0	11	4	7	0

#### A.7. Taxonomic distribution of reads assigned to functions in the Calvin-Benson-Bassham Cycle

Enzyme	Total Archaea	N. maritimus	Total Bacteria	Unclassified Bacteria	Proteobacteria	Other bacterial phyla
glyceraldehyde-3-phosphate dehydrogenase (NAD(P)) (EC:1.2.1.59)	34	25	30	24	0	6
transketolase (EC:2.2.1.1)	87	26	1402	929	299	174
phosphoribulokinase (EC:2.7.1.19)	0	0	45	28	16	1

Enzyme	Total Archaea	N. maritimus	Total Bacteria	Unclassified Bacteria	Proteobacteria	Other bacterial phyla
phosphoglycerate kinase (EC:2.7.2.3)	76	13	529	285	156	88
sedoheptulose-bisphosphatase (EC:3.1.3.37)	0	0	0	0	0	0
ribulose-bisphosphate carboxylase large chain (EC:4.1.1.39)	46	0	78	41	21	16
ribulose-bisphosphate carboxylase small chain (EC:4.1.1.39)	0	0	10	2	8	0
fructose-bisphosphate aldolase, class I (EC:4.1.2.13)	61	15	313	245	56	12
fructose-bisphosphate aldolase, class III (EC:4.1.2.13)	0	0	348	217	72	59
ribose 5-phosphate isomerase A (EC:5.3.1.6)	29	24	98	36	57	5
ribose 5-phosphate isomerase B (EC:5.3.1.6)	0	0	285	212	7	66
fructose-1,6-bisphosphatase II (EC:3.1.3.11)	0	0	139	70	49	20
fructose-1,6-bisphosphatase I (EC:3.1.3.11)	7	0	115	59	39	17
fructose-1,6-bisphosphatase III (EC:3.1.3.11)	0	0	5	0	5	0
glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating) (EC:1.2.1.13)	0	0	0	0	0	0
fructose-1,6-bisphosphatase II / sedoheptulose-1,7-bisphosphatase (EC:3.1.3.11 3.1.3.37)	0	0	0	0	0	0