Supplemental Information

Metagenomic and metatranscriptomic analyses reveal structure and dynamics of a dechlorinating community containing *Dehalococcoides mccartyi* and corrinoid-providing microorganisms under cobalamin-limited condition

Running title: High-throughput sequencing of dechlorinating communities

Yujie Men[1,†,‡,*], Ke Yu[1,†,‡], Jacob Bælum[2,†], Ying Gao[1,3], Julien Tremblay[4], Emmanuel Prestat[2,†], Ben Stenuit[1,†], Susannah G. Tringe[5], Janet Jansson[6], Tong Zhang[7], Lisa Alvarez-Cohen[1,2,*]

[1]Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA

[2]Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[3]School of Environment, Tsinghua University, Beijing, China

[4]National Research Council Canada, Montreal, QC, Canada

[5]DOE Joint Genome Institute, Walnut Creek, CA, USA

[6]Pacific Northwest National Laboratory, Richland, WA, USA

[7]The University of Hong Kong, Hong Kong

[*]Address correspondence to Yujie Men

3209 Newmark Civil Engineering Laboratory

205 N Mathews Ave, Urbana IL 61801

Email: ymen2@illinois.edu

Phone: (217) 244-8259

[†] Current addresses:

Yujie Men, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801;

Ke Yu, Peking University at Shenzhen, China;

Jacob Bælum, Chr Hansen A/S, Hørsholm, Denmark;

Emmanuel Prestat, Biostatistics & Bioinformatics at QIAGEN, Marseille Area, France;

Ben Stenuit, Earth and Life Institute, Catholic University of Louvain, Louvain-la-Neuve, Belgium;

‡ The two authors contributed equally to this study.

Supplemental methods

1. Library preparation for RNA sequencing

Structural RNA removal and mRNA enrichment

The purified total RNA was subjected to first strand cDNA synthesis using biotinylated universal 16S and 23S rRNA probes (1) and SuperScript® II reverse transcription (RT) system (Invitrogen, Life Technologies, Grand Island, NY) according to the manufacturer's instructions. A mixture of 2 µL dNTP (10 mM each), 2 µL biotinylated universal 16S primer mix (2.5 mM each), 2 µL biotinylated universal 23S primer mix (2.5 mM each), 1 µL SUPERaseIn RNase inhibitor (20U/µL), together with 8 µL RNA sample (100 ng total) was denatured at 65 °C for 3 min and cooled on ice for 2 min. Then, the reaction mix was amended with 4 µL 5X First-Strand buffer and 1 µL (200 U) SuperScript II reverse transcriptase and incubated at 42 °C for 50 min.) Dynabeads® MyOne™ Streptavidin C1 magnetic beads (Invitrogen) were then used to capture the structural RNA from the above RT mix. (A 1.5-mL microcentrifuge tube containing 20 µL suspension of beads was put on a magnetic stand for 1 min, the supernatant was then discarded. The beads were washed by 20 µL of solution A (DEPC-treated 0.1 M NaOH and DEPC-treated 0.05 M NaCl, filter sterilized), B (DEPC-treated 0.1 M NaCl, filter sterilized) and 1X First-Strand buffer, respectively. The washed beads were resuspended into 10 µL 1X First-Strand buffer.) A mixture of 5 µL washed beads were added into the 20 µL RT mix, and incubated at room temperature for 15 min. The beads were captured on a magnetic stand and the supernatant was transfered into another tube, with another 5 µL of washed beads added for a 30 min-incubation at room temperature. The beads were captured and the supernatant was transferred into a new tube. The collected beads were resuspended in 10 µL 1X First-Strand buffer, and the

supernatant was separated from the beads and combined with the previous ones, resulting in a total of 40 µL enriched mRNA solution.

Double stranded cDNA synthesis for linear amplification of mRNA

First stranded cDNA was first synthesized using random hexamer. 1µL primer (1µg/µL), 1µL dNTPs (10 mM each), and 1 µL (200 U) of SuperScript II reverse transcriptase were added into the 40 µL enriched mRNA solution and hybridized at 25 °C for 2 min, followed by incubation at 42 °C for 50 min. The reaction was then inactivated at 70 °C for 10 min.

The second strand was synthesized using NEBNext mRNA Second Strand Synthesis Module (New England Biolabs, Inc., Ipswich, WA) according to the manufacturer's instructions. The synthesized double stranded cDNA was purified and eluted into 10 µL nuclease-free water using MinElute PCR purification kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions.

Linear amplification

T7 amplification requires addition of the T7 promoter to DNA fragments through the ligation of a T7-incorporated adapter. This enables T7 RNA polymerase-mediated transcription in an in vitro reaction, which amplifies the template in a linear rather than an exponential manner (2). The resulting RNA fragments are subsequently converted to cDNA, in which the representation of different DNA fragments is highly similar to that of the original samples (2). The entire procedure includes steps of end repair, "A" tailing, adapter ligation, size selection, linear amplification, and final double stranded cDNA synthesis. (The details of each step are described in supplementary information).

The purified and quantified ds cDNA was blunted and phosphorylated using End-It DNA End-repair Kit (Epicentre, Madison, WI) according to the manufacturer's instructions. An

adenine was added to the 3'-end of DNA fragments using Klenow fragment (New England Biolabs) according to the manufacturer's instructions. The adapter A (containing P5 primer) and adapter B (containing T7 promoter and P7 primer) for Illumina sequencing were then ligated onto the DNA fragments using Ligalast rapid DNA ligation system (Promega, Madison, WI) according to the manufacturer's instructions). The 400-500 bp DNA fragments (including the adapters) shown on the E-gel system (Invitrogen) were excised and purified using QIAquick Gel Extraction Kit following the manufacturer's instructions. The adapter-ligated and size-selected DNA fragments were then linearly transcribed into RNA using MEGA short script T7 kit (Ambion, Life Technologies, Grand Island, NY) according to the manufacturer's instructions. The transcribed RNA was purified using RNeasy MinElute cleanup kit (Qiagen) following the manufacturer's instructions, and then reverse transcribed into double stranded cDNA using P5 primer (first strand cDNA synthesis) and P7 (second strand cDNA synthesis) primer according to the protocols described by others (2).

2. High throughput sequencing data analysis

The high throughput sequencing data were analyzed according to steps shown in Fig. S2. The following describes details of the major steps.

De novo assembly and ORF prediction

The raw DNA reads were first trimmed to 100 bp and filtered using criteria that all of the nucleotides in one read have a quality score equal or above 20. The trimmed and quality filtered reads from both HiTCEB12 and HiTCE samples were then submitted for de novo assembly by CLC genomics workbench version 6 (CLC bio, Boston, MA), using a kmer value of 63 and a minimum contig length of 1 kbp. The assembled contigs were submitted to Metagenmark server (3) for ORF/gene prediction.

Bin-genome construction and annotation

The difference in enrichment conditions of HiTCEB12 (with exogenous cobalamin) and HiTCE (without exogenous cobalamin) may result in shifts of microbial population abundances. A sequence-composition-independent, differential coverage binning method developed by others (4) was modified and applied to reconstruct the bin-genomes from the metagenomic datasets. Briefly, the coverage of each contig was calculated by mapping the reads after quality control against the contig sequence with a minimum similarity of 95% over 100% of the read length. Binning contigs into microbial population genomes were carried out by first plotting the coverage estimates of the two enrichments against each other for all contigs. The contigs clustered by differential coverage binning were further binned by tetranucleotide frequency and then extracted as bin-genomes.

The taxonomic assignment of the predicted ORFs was made by MEGAN5 (5) using the output file of BLASTx against the NCBI non-retardant protein sequence database (nr). The functional annotation of the predicted ORFs was assigned according to SEED subsystems database through MG-RAST, with a cut-off e-value of 10^{-5}.

Metatranscriptomic analysis

A rapid rRNA filter using DUK, an efficient kmer matching tool developed by JGI (https://sourceforge.net/projects/duk/) was applied to separate rRNA reads from the metatranscriptomic sequence pool. The non-rRNA reads were analyzed using modified Tophat and Cufflinks packages (6). In general, the RNA fragments were mapped against the predicted ORF/gene sequences. RPKM (Reads Per Kilobase pair transcript per Million total reads mapped) values were calculated using Cuffdiff, a part of Cufflinks package. Transcripts detected under both conditions with 5-fold difference in RPKM values were considered to be significantly differentially expressed, and were subjected to subsequent analyses. In order to investigate the gene regulation at a single-cell level, RPKM values normalized to one specific bin-genome (i.e.

reads per kilobase pair transcript per million reads mapped in one bin-genome) were also calculated using Bowtie2 alignment.

Supplemental Figures and Tables

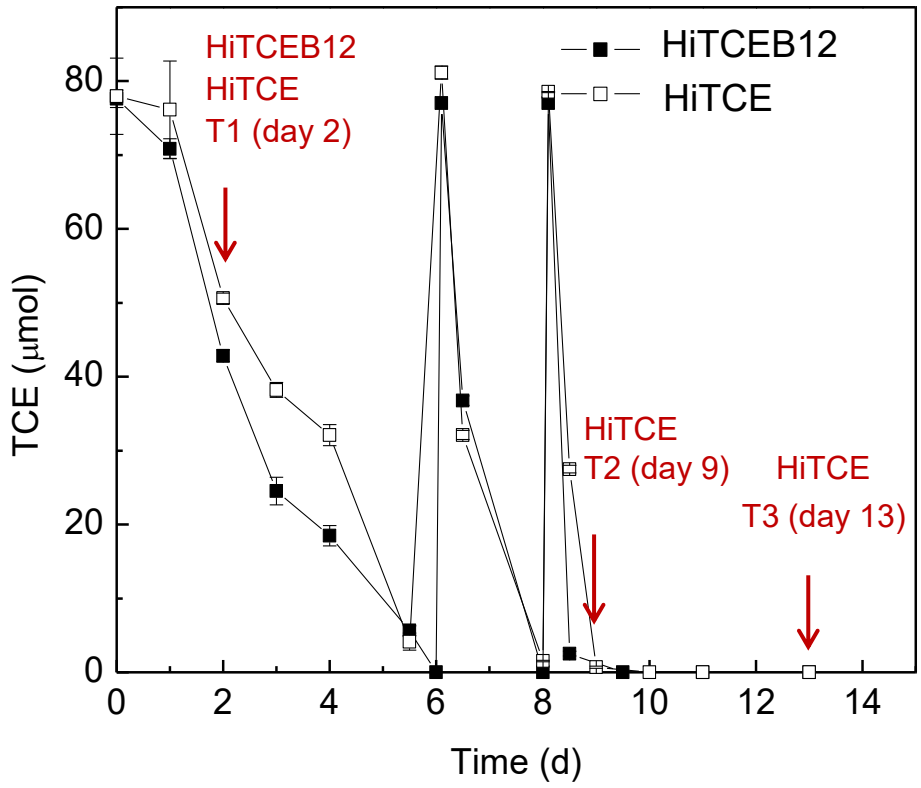

Figure S1. TCE degradation during a 13-day incubation (TCE was re-amended on day 6 and 8; red arrows indicate the time points at which samples were taken for metatranscriptomic sequencing)

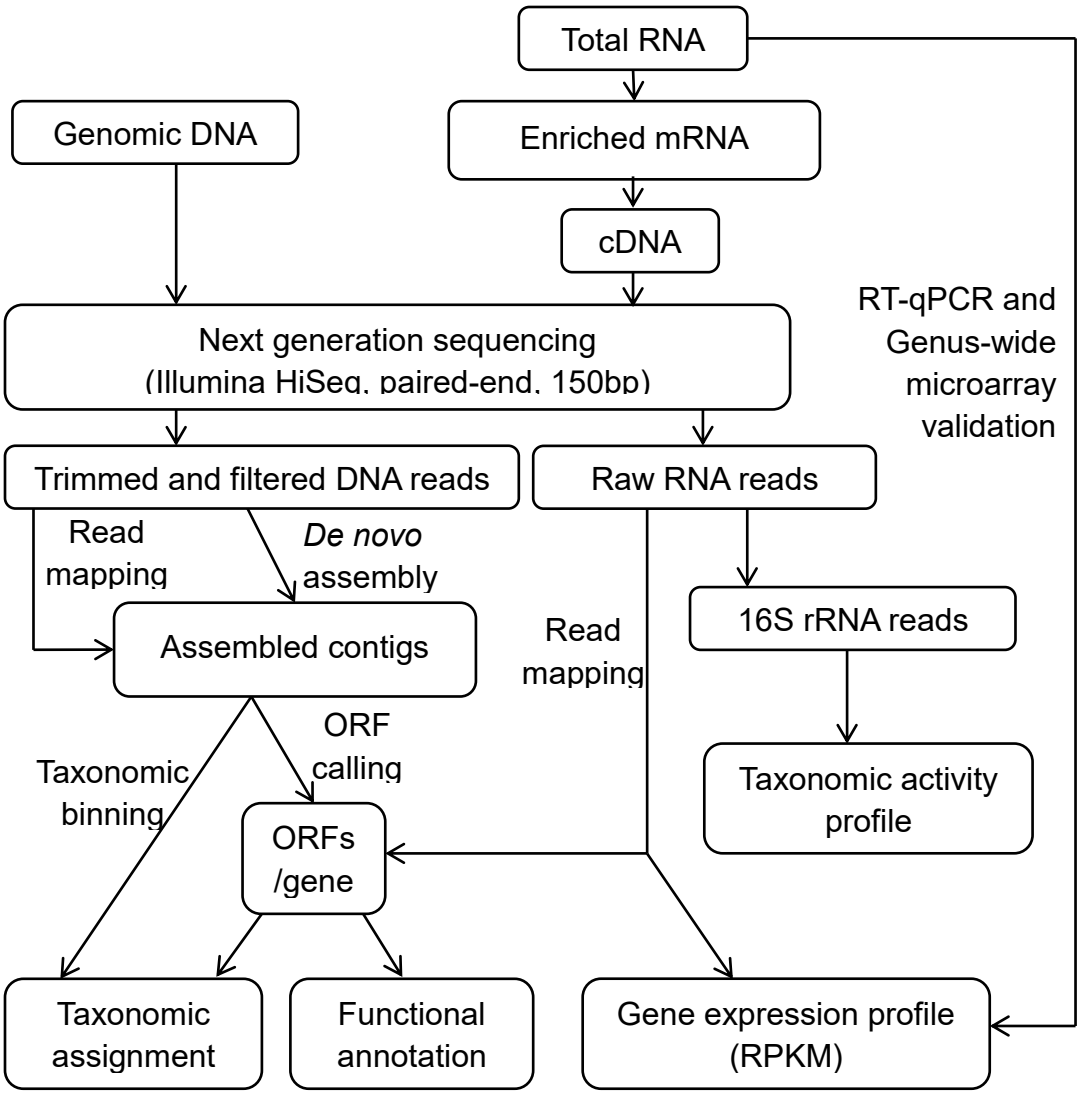

Figure S2. Schematic workflow of metagenomic and metatranscriptomic analysis. (ORF: open reading frame; RPKM: reads per kilobase pair of transcript per million reads mapped)

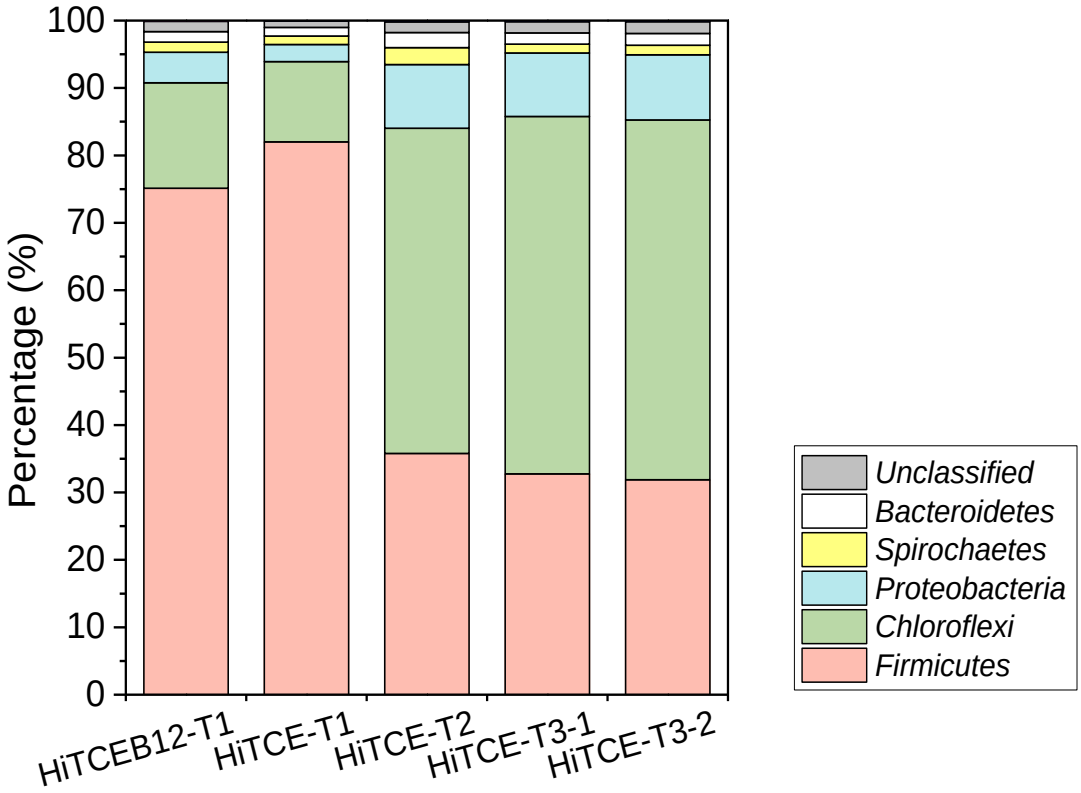

Figure S3. 16S rRNA distribution (at Phylum level) in metatranscriptomes (HiTCE-T3-1 and HiTCE-T3-2 are replicates).

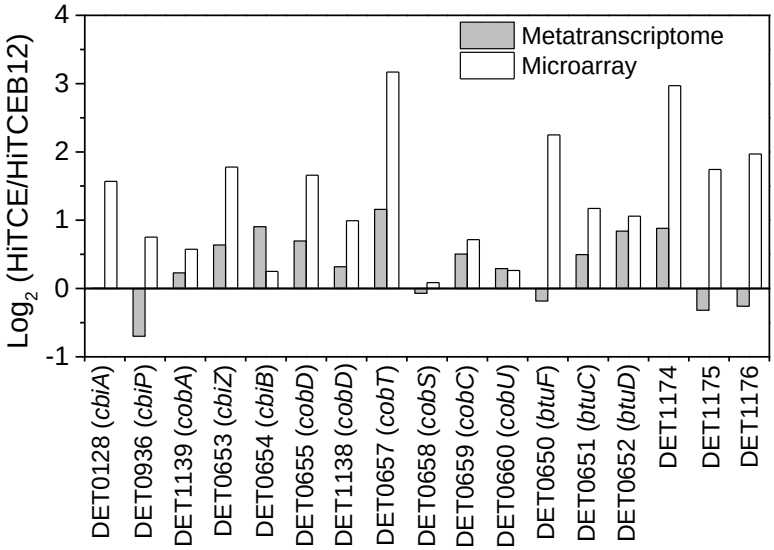

Figure S4. Differential expression of B_{12}-related genes in *Dehalococcoides* bin-genome at T1 by metatranscriptome and *Dehalococcoides* genus-wide microarray.

Table S1 Quantitative PCR primers targeting genes involved in cobalamin biosynthesis

Cobalamin biosynthesis enzymes		Gene ID	Primers (5'→ 3')
CobA	uroporphyrin-III C-methyltransferase	947	F440: AAGATCCAACCAAGGGCGAG R571: CCGGACGGCCATTTTCAATC
CysG	precorrin-2 dehydrogenase/siroheme synthase	944	F282: TGGGCGGGGTTTGTTATTCA R392: TTGCCGGCTGTAGAAACAGT
CbiK	cobalt chelatase	1227	F376: ACGGGGCAAGAAAATCAACC R519: GAGTGCTGCATAGGCGGTAT
CbiL	precorrin-2 C20-methyltransferase	6488	F483: TCCAGAAGAGTTGCGTGTCA R667: CACCGCGGAGACTGTCTAAA
CbiH	precorrin-3B C17-methyltransferase	6491	F176: TTGCTGTGGTTTCAAGTGGC R308: GATGCTCCTACCGCACTGAT
CbiF	precorrin-4 C11-methyltransferase	6489	F129: ATTTGCCAAACAGGGTGCTG R254: CTCGGGTCTCCCGTATGAAC
CbiG	cobalamin biosynthesis protein	6490	F540: TGTTAGAACCGCGACTGAGC R654: CGAATCCACGGTCAGCTCTT
CbiD	cobalt-precorrin-6A synthase	6485	F735: ATACAACCTTCCCGCTGAGG R906: CCTGGCATCTGCCATACGAT
CbiJ	precorrin-6x reductase	6492	F531: CGCCATGCAAGGTCCATTTT R636: ATCACTACCGCCAAGGTTGC
CbiT	precorrin-6B (C15)-methyltransferase	6487	F207: AGAAGAAGGCGTTCGCTTGA R314: GCTTGCGGTAAATCGCTCAG
CbiE	precorrin-7 (C5)-methyltransferase	N.D.	N/A
CbiC	Precorrin-8x methylmutase	6493	F176: AGGCTGGATGCCAAGCTTTA R307: TGCCAGGGTCTGCGATTAAG
CbiA	cobyrinic acid a,c-diamide synthase	6494	F591: TACAGAGCATGTTGCCGAGG R697: ATAAAGGCGGGGCAGTTTGA
CobA	cob(I)alamin adenosyltransferase	5467	F306: ACGCATTCCCCAACGGTATT R423: GGTACGCTTGAGGGTGACAA
CbiP	cobyric acid synthase	6496	F1308: TGGCTTGGTAATGGGCACTT R1416: TGTGCCGAGTGGTGTTAGAC
CbiB/ CobD	cobalamin biosynthesis protein	6497	F244: ATCGCAGTATGGGTTGGTGG R370: CCTTATAGCGTGCCTGCTCA
CobU	adenosylcobinamide-phosphate guanylyltransferase	6495	F406: GGCATGAGCATCGTACCTGA R526: CTGGGATTCCGCTGACAACT
CobT	nicotinate-nucleotide dimethylbenzimidazole-P phophoribosyl transferase	5484	F659: GCTGTGTTGGCGGTTTTGAT R787: CATTGATCTTGGCCGCACAC
CobC	alpha-ribazole-5'-P phosphatase	6501	F171: TAGAGCCTACCATACGGCGG R269: CCTTCCCATTCCCCAAAGGA
CobS_1	cobalamin 5'-phosphate synthase	5485	F86: GTTGGTCACCGGAGTCCTTT R211: GGACGGGTACATCATTGCCA
CobS_2	cobalamin 5'-phosphate synthase	6498	F463: TTAGGCCGCTTTGCTATGGT R635: ACCGCAATACAAATGGCAGC
CbiZ	adenosylcobinamide amidohydrolase	1215	F710: ACGGCTTATCGCCCAATCAA R879: GGTCACTAAGGCAGGCAGTT
RpoB	RNA polymerase beta subunit	103	F1775: TCTCGCCTAAGCAGGTCGTA R1911: AACAATTGGAGCCTGCGTTC
TceA	Reductive dehalogenase	9272	F: ATCCAGATTATGACCCTGGTGAA R: GCGGCATATATTAGGGCATCTT

N.D.: not detected; N/A: not available

Table S3 Presence of anaerobic cobalamin biosynthesis genes in bingenomes and the corresponding transcripts in the metatranscriptomes

Gene	Corresponding enzymes	Bingenomes			
		Veillonellaceae	*Desulfovibrio*	*Sedimentibacter*	*Dehalococcoides*
		(presence of gene/transcript)			
cobA	uroporphyrin-III C-methyltransferase	+/-	+/-	+/-	-
cysG	precorrin-2 dehydrogenase/siroheme synthase	**+/+**	+/-	+/-	-
cbiK	cobalt chelatase	+/-	+/-	+/-	-
cbiL	precorrin-2 C20-methyltransferase	+/-	+/-	+/-	-
cbiH	precorrin-3B C17-methyltransferase	+/-	+/-	+/-	-
cbiF	precorrin-4 C11-methyltransferase	+/-	+/-	+/-	-
cbiG	cobalamin biosynthesis protein	+/-	+/-	+/-	-
cbiD	cobalt-precorrin-6A synthase	+/-	+/-	+/-	-
cbiJ	precorrin-6x reductase	+/-	-	+/-	-
cbiT	precorrin-6B (C15)-methyltransferase	+/-	-	-	-
cbiE	precorrin-7 (C5)-methyltransferase	-	+/-	+/-	-
cbiC	Precorrin-8x methylmutase	+/-	+/-	+/-	-
cbiA	cobyrinic acid a,c-diamide synthase	+/-	+/-	+/-	**+/+**
cobA	cob(I)alamin adenosyltransferase	+/-	+/-	+/-	**+/+**
cbiP	cobyric acid synthase	+/-	+/-	+/-	**+/+**
cbiB/ *cobD*	cobalamin biosynthesis protein	+/-	+/-	+/-	**+/+**
cobU	adenosylcobinamide-phosphate guanylyltransferase	+/-	+/-	+/-	**+/+**
cobT	nicotinate-nucleotide dimethylbenzimidazole-P phophoribosyl transferase	+/-	-	-	**+/+**
cobC	alpha-ribazole-5'-P phosphatase	+/-	+/-	+/-	**+/+**
cobS	cobalamin 5'-phosphate synthase	**+/+***	+/-	+/-	**+/+**
cbiZ	adenosylcobinamide amidohydrolase	+/-	-	-	**+/+**

*Transcripts of one out of two *cobS* genes were detected in the metatranscriptome.

References

1. Stewart FJ, Ottesen EA, DeLong EF. 2010. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. ISME J 4:896-907.

2. Hoeijmakers WA, Bartfai R, Francoijs KJ, Stunnenberg HG. 2011. Linear amplification for deep sequencing. Nat Protoc 6:1026-1036.

3. Zhu W, Lomsadze A, Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res 38:e132.

4. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol 31:533-538.

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

6. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562-578.