# Supplementary information

## Direct utilization of polysaccharides for methane production by a single deep-sea methanogen

Rikuan Zheng[1,2,3#], Chong Wang[1,2,3#], Chaomin Sun[1,2,3,4*]

[1]CAS and Shandong Province Key Laboratory of Experimental Marine Biology & Center of Deep Sea Research, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China

[2]Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology Center, Qingdao, China

[3]Center of Ocean Mega-Science, Chinese Academy of Sciences, Qingdao, China

[4]College of Earth Science, University of Chinese Academy of Sciences, Beijing, China

[#] These authors contributed equally to this work

[*] Corresponding author

Chaomin Sun Tel.: +86 532 82898857; fax: +86 532 82898857.

E-mail address: sunchaomin@qdio.ac.cn

**Key words:** Polysaccharide, polyphosphate, deep sea, methanogenesis, carbon cycle

## Supplementary Methods

**Transcriptional profiling of strain ZRKC1 cultured in basal medium supplemented with 0.01 g/L yeast extract alone, or in combination with 2.0 g/L raffinose, chitin, or *Tremella* polysaccharide**

**(1) Library preparation for strand-specific transcriptome sequencing.** A total amount of 3 μg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (NEB, USA) according to manufacturer's recommendations and index codes were added to attribute sequences to each sample. The rRNA is removed using a specialized kit that leaves the mRNA. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5×). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNaseH$^-$). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. In the reaction buffer, dNTPs with dTTP were replaced by dUTP. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop structure was ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 150~200 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 μL USER Enzyme (NEB，USA) was used with size-selected, adaptor-ligated cDNA at 37 °C for 15 minutes followed by 5 minutes at 95 °C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system.

**(2) Clustering and sequencing.** The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit

v3-cBot-HS (Illumia) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina Hiseq platform and paired-end reads were generated.

**(3) Data analysis.** Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing ploy-N and low quality reads from raw data. At the same time, Q20, Q30 and GC content the clean data were calculated. All the downstream analyses were based on the clean data with high quality. Reference genome and gene model annotation files were downloaded from genome website directly. Both building index of reference genome and aligning clean reads to reference genome were used Bowtie2-2.2.3[1]. HTSeq v0.6.1 was used to count the reads numbers mapped to each gene. And then FPKM of each gene was calculated based on the length of the gene and reads count mapped to this gene. FPKM, expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced, considers the effect of sequencing depth and gene length for the reads count at the same time, and is currently the most commonly used method for estimating gene expression levels[2].

**(4) Differential expression analysis.** Differential expression analysis of two conditions/groups (two biological replicates per condition) was performed using the DESeq R package (1.20.0)[3]. DESeq provide statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting *P*-values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted $P < 0.05$ found by DESeq were assigned as differentially expressed. (For DEGSeq without biological replicates) Prior to differential gene expression analysis, for each sequenced library, the read counts were adjusted by edgeR program package through one scaling normalized factor. Corrected *P*-value of 0.005 and $\log_2$ (Fold change) of 1 were set as the threshold for significantly differential expression.

**(5) GO and KEGG enrichment analysis of differentially expressed genes.** Gene Ontology (GO) enrichment analysis of differentially expressed genes was implemented by the GOseq R package, in which gene length bias was corrected[4]. GO terms with corrected *P* value less than 0.05 were considered significantly enriched by differential expressed genes. KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (http://www.genome.jp/kegg/). We used KOBAS software to test the statistical enrichment of differential expression genes in KEGG pathways.

## References


1 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359, (2012).

2 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106, (2010).

3 Wang, L. K., Feng, Z. X., Wang, X., Wang, X. W. & Zhang, X. G. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136-138, (2010).

4 Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14, (2010).

# Supplementary Table list

***Supplementary Tables are provided as separate excel files. Supplementary tables are listed as follows:***

**Supplementary Table 1.** Analysis of pectin degrading enzymes in deep-sea methanogenic archaea.

**Supplementary Table 2.** Analysis of the isolated sources of cultured archaea.

**Supplementary Table 3.** Primers used for qRT-PCR.

# Supplementary Figure Legends

**Extended Data Fig. 1 | Isolation and phylogenetic characterization of eight deep-sea methanogenic archaeal strains. a**, Diagrammatic scheme of enrichment and isolation of methanogenic archaea. **b**, Phylogenetic analysis of strains ZRKC1, ZRKC2, ZRKC3, ZRKC4, ZRKC5, WCC1, WCC4, and WCC5 based on full-length 16S rRNA gene sequences. The *Methanofollis aquaemaris* N2F9704 were used as the outgroup. The tree is inferred and reconstructed by maximum likelihood, with nodes exhibiting >80% bootstrap support denoted by gray circles (expressed as percentages of 1000 replications). Scale bar, 0.1 substitutions per nucleotide position.

**Extended Data Fig. 2 | Quantification of methanol in sterilized culture media.** "Medium" indicates the basal medium alone; "Medium+3.0 g/L methyl ester pectin" indicates the basal medium supplemented with 3.0 g/L methyl ester pectin; "Medium+10 μL methyl formate" indicates the basal medium supplemented with 10 μL methyl formate. All media were sterilized by autoclaving at 115 °C for 30 min, consistent with previous experiments.

**Extended Data Fig. 3 | The potential methyl esterotrophic methanogenesis compared to known methanogenic pathways.** Black labels represent methanogenic substrates, while red labels indicate the types of methanogenesis. A putative methyl esterotrophic pathway is proposed (orange line), suggesting a direct role of methyl esters in methane production.

**Extended Data Fig. 4 | Growth and methane production of the reference strain 8220 in the modified DSMZ 120 medium supplemented with different concentrations of the phosphatase inhibitor.** Error bars show mean ± SD. "ns" denotes no significant difference ($P > 0.05$).

**Extended Data Fig. 5 | Growth of strain ZRKC1 in seawater medium or *in situ* seawater medium, both supplemented with methanol.**

**Extended Data Fig. 6 | TEM observation of polyP in eight deep-sea methanogenic strains (ZRKC1, ZRKC2, ZRKC3, ZRKC4, ZRKC5, WCC1, WCC4, WCC5).** The black granules on the cell surface are identified as polyP.

**Extended Data Fig. 7 | RT-qPCR analysis the expression of two polyphosphate kinase genes (*ppk1* and *ppk2*) during the growth of strain ZRKC1.**

**Extended Data Fig. 8 | TEM observation of polyP granules in *E. coli* BL21. a**, TEM image of *E. coli* BL21 cells transformed with the empty vector pET28a. **b**, TEM images of polyP granules in *E. coli* BL21 cells overexpressing PPK1. **c**, TEM images of polyP granules in *E. coli* BL21 cells overexpressing PPK2. **d**, TEM image of polyP granules in *E. coli* BL21 cells overexpressing PPK1, supplemented with 8 mM polyphosphate kinase inhibitor (mesalamine). Yellow arrows indicate the location of polyP granules.

**Extended Data Fig. 9 | *In vitro* analysis of the key role of PPK1 in directing polyP formation. a**, SDS-PAGE analysis of purified PPK1 and PPK2 proteins. Lane 1: protein markers (sizes in kDa indicated on the left); lane 2: purified PPK1; lane 3: purified PPK2. **b**, *In vitro* activity assays of purified PPK1 and PPK2. Specific enzyme activity is expressed as U/mg protein, where one unit corresponds to the enzyme amount that converts 1 μmol of substrate per minute. **c**, Predicted structure of PPK1 bound to $Mg^{2+}$, as determined by the AlphaFold server. The blue sphere represents $Mg^{2+}$. **d-g**, Effect of varying $Mg^{2+}$ concentrations (0, 10, 50 mM) on inorganic phosphate levels (d), polyP levels (e), growth (f), and methane production (g) during ZRKC1 growth.

**Extended Data Fig. 10 | PPK1 homologs are widely distributed among archaea across various habitats.** Phylogenetic analysis was conducted based on the amino acid sequences of PPK1. The tree was constructed using maximum likelihood, with nodes showing >80% bootstrap support indicated by light blue circles (representing percentages from 1,000 replications). The scale bar represents 1.0 substitution per nucleotide position. Shapes in the tree are color-coded according to the habitat categories of the corresponding archaea. The analysis identifies distinct clades, labeled as Clade I, Clade II, Clade III, Clade IV, Clade V, and others.