Supplementary Material: Lanthanide-Dependent Methylotrophs of the Family Beijerinckiaceae: Physiological and Genomic Insights

Carl-Eric Wegner[1#], Linda Gorniak[1], Stefan Riedel[1], Martin Westermann[2], and Kirsten Küsel[1,3]

[1]Institute of Biodiversity, Aquatic Geomicrobiology, Friedrich Schiller University, Dornburger Str. 159, 07743 Jena, Germany

[2]Electron Microscopy Centre, University Hospital Jena, Ziegelmühlenweg 1, 07743 Jena, Germany

[3]German Center for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5E, 04103 Leipzig, Germany

Corresponding author:

Carl-Eric Wegner[1]

Email address: carl-eric.wegner@uni-jena.de

SUPPLEMENTAL MATERIAL

Supplementary Information. Complementary information about used materials and methods.

FIG S1. Macroscopic imaging of the three methylotrophic strains isolated from early-industrial soft coal slags. RH CH11 (A), RH AL1 (B), RH AL8 (C); bar = 1 mm.

FIG S2. Scatterplot of 16S rRNA gene identities versus percentage of conserved proteins (POCP). We performed pairwise comparisons based on the isolate genomes and available Beijerinckiaceae genomes and determined 16S rRNA gene identities and POCP values. The color code refers to comparisons between different genera and within the same genus. The lower dotted line indicates the 50% POCP threshold suggested by Qin et al., 2014 (1) for genus-level delineation, the upper dotted line indicates the POCP threshold that we found to be robust for genus-level delineation within the Beijerinckiaceae.

FIG S3. Functional gene PCR targeting genes linked to C_1-metabolism. Details about PCR conditions and used primers can be found in the Supplementary Information. *mmoX* PCR products obtained for the primer pair mc2 were checked by Sanger-sequencing and turned out to be unspecific. *xoxF1* PCR products were not in the expected size range and not further considered. The same was true for RH AL1 in case of *mauA*.

FIG S4. Reverse-transcription PCR of *xoxF*. Total RNA from exponentially growing cultures was extracted and reverse-transcribed. Yielded cDNA was used as template

for PCR using available *xoxF* primers (2). Positive PCR products for *xoxF4* although yielded isolates lack *xoxF4* genes are a consequence of the inherent cross-specificity of the *xoxF4* primer set for *xoxF5*.

FIG S5. Phylogenetic treeing of PQQ ADH sequences. Hit sequences (TABLE S5) were extracted and subjected to treeing using the constructed PQQ ADH sequence database. Sequences were aligned using muscle (v. 3.8.31) (3) and phylogenetic treeing was done with mega (v. X) (4) using the Neighbor-Joining method (5) and validated by bootstrapping (6). Clades highlighted in grey (e.g. XoxF5) are supposed to rely on REE as co-factors. Bootstrap support is indicated by colored circles placed on the respective node.

FIG S6. Results of screening Beijerinckiaceae genomes for PQQ ADHs using pHMMs. Squares indicate the number of identified genes encoding the respective enzyme. NCBI assembly accessions are given in parentheses.

FIG S7. Sequence analysis of the lanmodulin protein encoded by Beijerinckiacea bacterium RHAL1. Calmodulin is shown as representative EF-hand containing protein. Residues in the EF-hand motifs that contribute to metal and lanthanide-selectivity are bolded in orange in the lanmodulin of *M. extorquens* AM1 and the lanmodulin-homolog in Beijerinckiaceae bacterium RH AL1. Differences in RHAL1 with respect to these residues are shown in black and boxed with a red-dotted line. Accessions are given in parentheses.

FIG S8. Estimation of indels present in the assembly of RH AL1. Indels present in

the canu (7) assembly of RH AL1 were assessed by querying predicted amino acid sequences against NCBI nr with diamond (8) and plotting the length ratio of query and subject (top left). The same procedure was repeated after assembly polishing with racon (9) and pilon (10) (top right) and indel correction with pacbio-utilities https://github.com/douglasgscofield/PacBio-utilities) (bottom left).

TABLE S1. DNA distance matrix of partial and full-length 16S rRNA gene sequences of isolated soft coal slag methylotrophs and closest related Beijerinckiaceae. Closely related 16S rRNA gene sequences were identified based on partial 16S rRNA gene sequences from obtained isolates using SINA (11) and retrieved from SILVA (release 132) (12). Sequences are named according to the following scheme: accession number|length or coordinates in case of genome/fosmid sequences|coordinates of the aligned part.

TABLE S2. Results of blastn searches of partial and full-length 16S rRNA gene sequences from soft coal slag isolates against NCBI's non-redundant nucleotide collection and NCBI 16S ribosomal RNA sequences. In case of the later the best ten hits are reported based on bit score, for the former the five best hits are reported.

TABLE S3. Fraction of orthologous genes, 16S rRNA identities, and percentages of conserved proteins (POCP) determined based on pairwise comparisons of our isolate genomes and available Beijerinckiaceae genomes. The fraction of orthologous genes was determined with compareM as described in the Supplementary Information. 16S rRNA gene identities were determined based on extracted full-length 16S rRNA gene sequences. POCP values were calculated as

outlined in the Supplementary Information.

TABLE S4. Elemental analysis of soft coal slag material based on total and sequential digestion. Total and bioavailable lanthanides were determined based on total and sequential digestion followed by ICP-MS, respectively. We define the bioavailable portion of lanthanides here as fractions I (mobile) and II (exchangeable) according to the sequential digestion method of Zeien and Brümmer (13, 14). The nomenclature of the sampling sites is according to Wegner and Liesack (15). RH1 and RH2 refer to the two slag deposition sites from which sample material was collected for enrichment and isolation, while RHRef describes undisturbed forest soil from the surrounding of RH1 and RH2. MS = average of triplicated measurements, SD = standard deviation.

TABLE S5. HMM-based screening for PQQ ADHs. We constructed a PQQ ADH sequence database based on Keltjens et al. (16) and created clade-specific HMM profiles. Sequence subsets for individual clades (e.g. XoxF1) were extracted, aligned with muscle (v. 3.8.31) (3), and HMMs built with hmmbuild (v. 3.1b2). These profiles were then merged and used to query the amino acid sequences of coding sequences in the assembled genomes with hmmsearch (v. 3.1b2) (17). The best hits based on e-value and bit-score were kept for putative PQQ ADH sequences. Columns represent encoded candidate PQQ ADH proteins identified in the three isolate genomes. Rows are the bit-scores of the best hit from hmmsearch queries using the respective profile HMMs.

TABLE S6. Basic characteristics of the assembled and annotated soft coal slag

methylotroph genomes.

TABLE S7. Selected gene descriptions and expression values of the genes coding for proteins displayed in Figure 6 linked to C_1-metabolism, central carbohydrate metabolism, central energy metabolism, sulfur metabolism, nitrogen metabolism, polyhydroxybutyrate biosynthesis, polyphosphate biosynthesis, transporters, and motility. Average expression values are taken from Table S8. Log_2RPKM = log_2 reads per kilobase million

TABLE S8. Transcriptome data of Beijerinckiaceae bacterium RHAL1 grown on methanol (1%, [v/v]) as carbon source and with supplemented lanthanum (1 µM). A, B, C refer to data from three biological replicates. Gene expression is given in log_2RPKM. Expression values were averaged (AVG).

TABLE S9. Results of blastp queries of gene products linked to lanthanide-dependent metabolism in *M. extorquens* AM1 against Beijerinckiaceae bacterium RHAL1. With the exception of LutCD and XoxG, only hits matching an e-value threshold of 1e -3 are reported. Expression values are taken from Table S8.

NOTE: All supplementary tables are provided in one combined spreadsheet ("Supplemental File 2").

SUPPLEMENTARY INFORMATION

Isolation. Solid VL55 medium (18) was inoculated with 200 µL of the 10^{-4} to 10^{-6} dilutions of agitated slag slurries and incubated for up to 8 weeks in the dark at room

temperature. Per litre of solid medium, 15 g of gellan gum were added as solidifying agent. For proper solidification additional divalent cations were added to the basal medium in form of $MgSO_4 \times 7\ H_2O$. Pectin, xylan (0.05% [w/v]) or a mixture of arabinose, glucose, xylose and N-acetylglucosamine (0.5 mM each) were added as carbon source. Aluminium-mediated stress was mimicked by adding up to 5 mM aluminium in the form of $KAl(SO_4)_2 \times 12\ H_2O$. The pH of used media was adjusted to 5.2-5.5 or pH 4 in case of aluminium containing medium.

16S rRNA gene-targeting colony PCR. Cell material from grown colonies was subjected to colony PCR for 16S rRNA gene-based identification. Cell material was collected with sterile 10 µL pipette tips, resuspended in 20 µL double-distilled water and boiled at 95°C for 5 minutes. The resulting cell lysate was used as template material for 16S rRNA gene PCR. 50 µL reactions were set up: 5 µL 10 × high yield buffer (Jena Bioscience, Jena, Germany), 1 µL dNTPs (10 mM each) (Bioline, Luckenwalde, Germany) , 1 µL of both primers (10 pmol/µL each [27f/1392r]), 0.5 µL of Taq polymerase (5 U/µL) (Jena Bioscience), and 40.5 µL of PCR-grade water. PCR reactions were denatured at 95°C for 5 minutes, followed by 30 cycles of: denaturation (95°C, 60 seconds), annealing (55°C, 60 seconds), and elongation (72°C, 90 seconds), followed by a final elongation step of 5 minutes. PCR products were checked by agarose gelelectrophoresis and gel purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. Purified PCR products were Sanger-sequenced by Macrogen (Amsterdam, The Netherlands).

PCR genotyping of C_1-metabolism related genes. All PCR protocols included an

initial denaturation (94°, 5 minutes) and a final elongation (72°C, 5 to 10 minutes). *pmoA* gene amplicons were generated using the primer sets A189f/A650r (21), A189f/A682r (22), and A189f/mb661r (23). For all three assays, PCR amplifications were carried out under the following conditions (24): 30 cycles consisting of denaturation (94°C, 30 seconds), annealing (56°C, 60 seconds), and elongation (72°C, 60 seconds). Three different primer pairs were used to detect *mmoX*, mmoXA/mmoXmc1, mmoXA/mmXmc2, and mmoXmc3/mmoXmc1 (25). The PCR protocol was: 35 cycles of denaturation (94°C, 60 seconds), annealing (55°C, 60 seconds), and elongation (72°C, 60 seconds). The primer set F1003/R1561 (26) and the more recent, degenerate primer set F1003degen/R1561degen (27) were used to identify *mxaF*: 30 cycles of denaturation (94°C, 45 seconds), annealing (60°C, 60 seconds), elongation (72°C, 90 seconds) (27). *xoxF* was shown to be phylogenetically diverse comprising multiple subgroups (2). Five different primer pairs (xoxF1-f/xoxF1-r, xoxF2-f/xoxF2-r, xoxF3-f/xoxF3-r, xoxF4-f/xoxF4-r, and xoxF5-f/xoxF5-r) were used with two different touchdown thermal profiles (2) to check for their presence. Primer sets mauAf1/mauAr1 and gmaS-557f/gmaS-907r (28, 29) were used to amplify *mauA* (35 cycles, denaturation: 94°C, 60 seconds; annealing: 48°C, 60 seconds; elongation: 72°C, 60 seconds) (30) and *gmaS* (touchdown 10 cycles, denaturation: 94°C, 45 seconds; annealing: 60°C, 45 seconds [-1°C/cycle]; elongation: 72°C; 30 cycles, denaturation: 94°C, 45 seconds; annealing: 56°C, 45 seconds; elongation: 72°C, 60 seconds) (29).

Electron microscopy. Fixed samples were washed three times with cacodylate buffer, and post-fixed with 1% osmium tetroxide in cacodylate buffer for 2 hours at 20°C. Samples were subsequently dehydrated in an ascending ethanol series and

stained with 2% (w/v) uranyl acetate in 50% (v/v) ethanol. The samples were embedded in Araldite resin (Plano, Wetzlar, Germany) according to manufacturer's instructions. Ultrathin sections (70 nm thickness) were cut using an ultramicrotome Ultracut E (Reichert-Jung, Vienna, Austria) and mounted on Formvar-carbon coated 100 mesh grids (Quantifoil, Großlöbichau, Germany). The ultrathin sections were stained with lead nitrate for 10 minutes (31) and examined in a Zeiss CEM 902 A electron microscope (Carl Zeiss AG, Oberkochen, Germany) and imaged using a TVIPS 1k Fast-Scan CCD-Camera (TVIPS, Munich, Germany).

Genome sequencing and assembly. Genomic DNA was extracted using the NucleoSpin® Microbial DNA kit (Macherey-Nagel, Düren, Germany). Sequencing libraries for RH CH11 and RH AL8 were prepared with Nextera sample kits (Illumina, San Diego, CA, USA). Illumina sequencing (2 × 150 bp) was done on an Illumina HiSeq 3000 instrument. The library of RH AL1 for long-read sequencing was prepared using PacBio® SMRTBell reagents (Pacific Biosciences, Menlo Park, CA, USA) and sequenced on a PacBio RS II instrument. Library preparation and sequencing was done by GATC Biotech GmbH (Constance, Germany). Default settings were used for all tools during data processing if not stated otherwise. Illumina data was inspected with fastQC (http://www.bioinformatics.babraham.ac.uk/project/fastqc) and adapter-trimmed and quality-filtered with bbduk (settings: minlen=75 qtrim=rl trimq=20 ktrim=r k=25 mink=11) (v. 38.26) (32). Draft genomes of RH CH11 and RH AL8 were assembled with spades (v. 3.13.0) (33). Assembling the PacBio data (30,725 reads, N50 = 11,451 bp) with canu (v. 1.5) (7) recovered two contigs (4.24 Mb, 135 Kb). Mapping the two contigs against each other with nucmer (-maxmatch -nosimplify, show-

coords -lrcTH) (v. 3.1) (34) revealed that the RH AL1 genome comprises a circular chromosome and plasmid. The start position of the circular chromosome was fixed with circlator (v. 1.5.5) (35) so that it starts with *dnaA*. We identified potential indels in the assembly of RH AL1 by querying predicted protein sequences against NCBI nr using diamond (v. 0.9.19) (8) and plotting the ratio between query and subject length. These could not be resolved by generating a consensus sequence of PacBio reads with racon (v. 1.3.3) (9) and pilon (v. 1.23) (10). Considering the very close relatedness of our three isolates (100% sequence identity for their 16S rRNA genes, average nucleotide identity values ≥ 99.9, percentage of conserved proteins ≥ 98%), we felt confident to use available short read data from one of the other two strains to correct present indels. Short read data from RH AL8 and pacbio-utilities (https:github.com/douglasscofield/PacBio-utilities) resolved most of the indels (FIG S8). The assembly of RH AL1 was used for scaffolding the RH AL8 and RH CH11 genomes with ragout (v.2.2) (36).

Phylogenomics. Besides from Beijerinckiaceae genomes related to our isolated strains, we collected multiple Methylobacteriaceae and Methylocystaceae genomes using ncbi-genome-download (https://github.com/kblin/ncbi-genome-download) for being used as outgroup. Anvio (v.5.5) (37) contig databases were generated using the program "anvi-gen-contigs-database" (non-default settings: –skip-mindful-splitting). Next, genomes were annotated against Pfam (38) and NCBI COG (39) with the programs "anvi-run-ncbi-cogs" (–sensitive) and "anvi-run-pfams". An anvio genomes storage database was created using the program "anvi-gen-genomes-storage". The pangenome was calculated with "anvi-pan-genome" (–minbit 0.5 –mcl-inflation 6), which uses blastp (blast+, v.2.9.0) (20) for calculating similarities of each

amino acid sequence across genomes, the minbit heuristic first implemented in ITEP (40) to identify and remove weak amino acid sequence matches, and the MCL algorithm (41) to identify clusters based on amino acid sequence similarity searches. The output was visualized and single copy core genes binned using the interactive interface of anvio by calling the program "anvi-display-pan". The encoded amino acid sequences of single copy core genes were extracted, individually aligned, concatenated, and subjected to treeing as described in the manuscript. DNA distance matrices of full-length 16S rRNA gene sequences were done with dnadist (v.3.697) (42). The percentage of conserved proteins (POCP) (1) was determined by using compareM (v.0.0.23) (https://github.com/dparks1134/CompareM) (settings: 10 -5 e-value, 30% sequence identity, and ≥70% alignment length) to identify orthologous genes shared among Beijerinckiaceae genomes. compareM relies on diamond (v.0.09.22) (8) for homology searches and prodigal (v. 2.6.3) (43) for gene prediction. The POCP between Beijerinckiaceae genomes was subsequently calculated applying the formula, $[(2 \times S)/(T1+T2)] \times 100\%$. S is the number of genes shared between the two compared genomes, while T1 and T2 represent the number of encoded proteins in the two genomes (1, 44).

SUPPLEMENTARY FIGURES

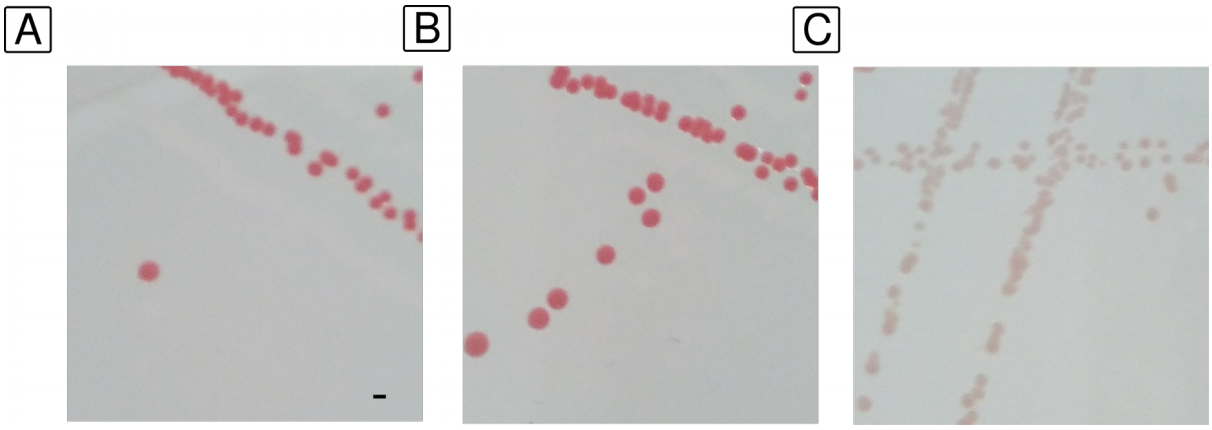

Fig S1.

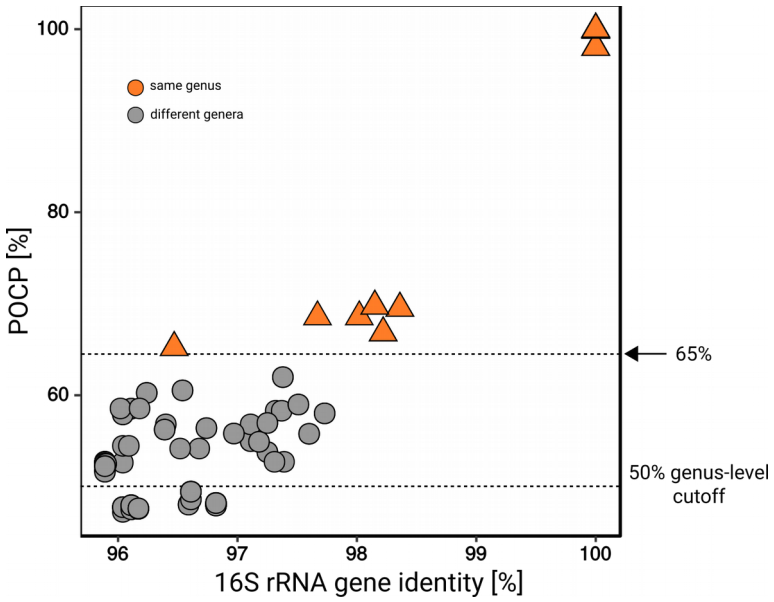

Fig S2.

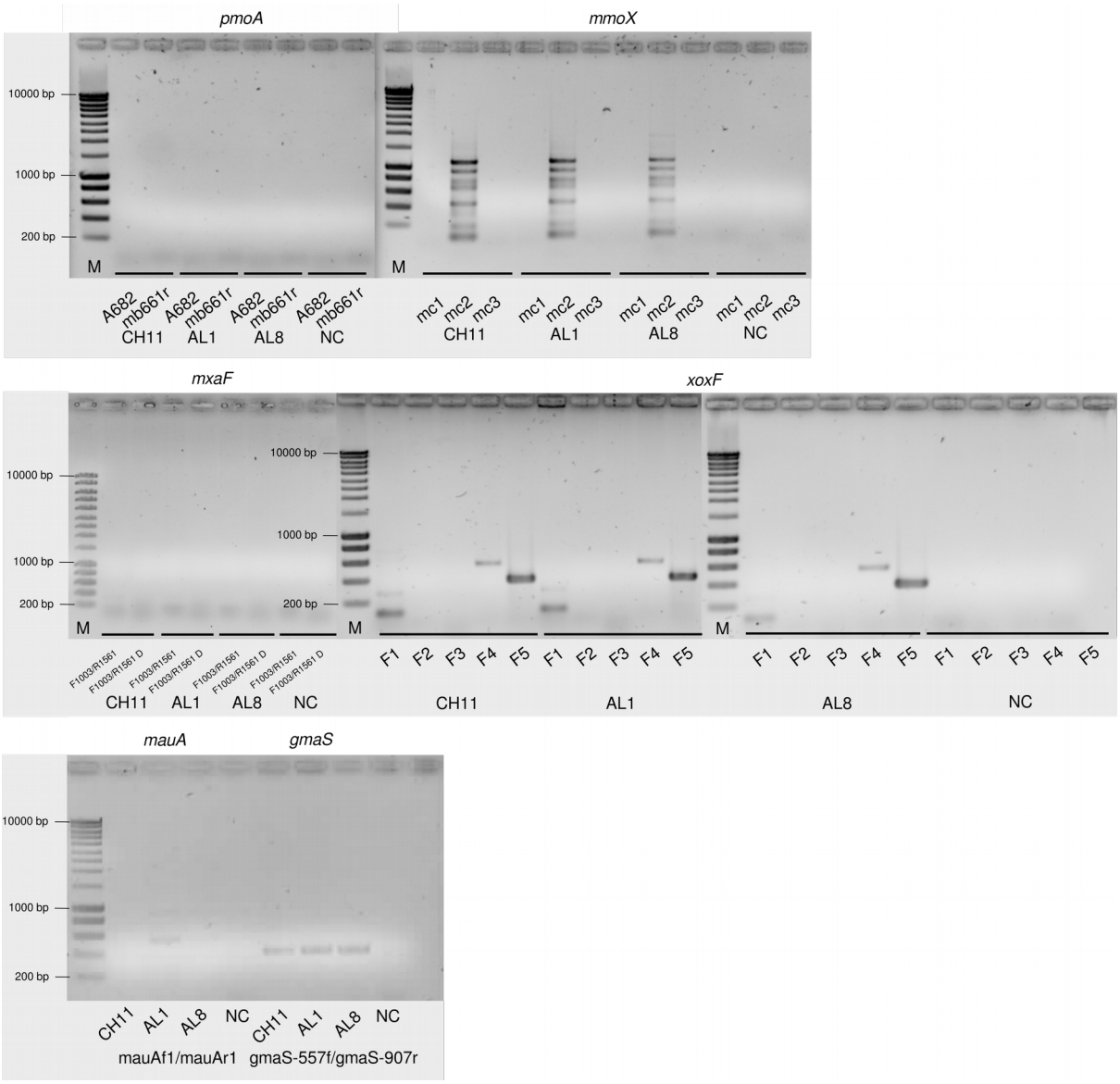
pmoA
mmoX
10000 bp
1000 bp
200 bp
M
A682
mb661r
CH11
AL1
AL8
NC
mc1
mc2
mc3
mxaF
xoxF
F1003/R1561
F1003/R1561 D
F1
F2
F3
F4
F5
mauA
gmaS
CH11
AL1
AL8
NC
mauAf1/mauAr1
gmaS-557f/gmaS-907r

Fig S3.

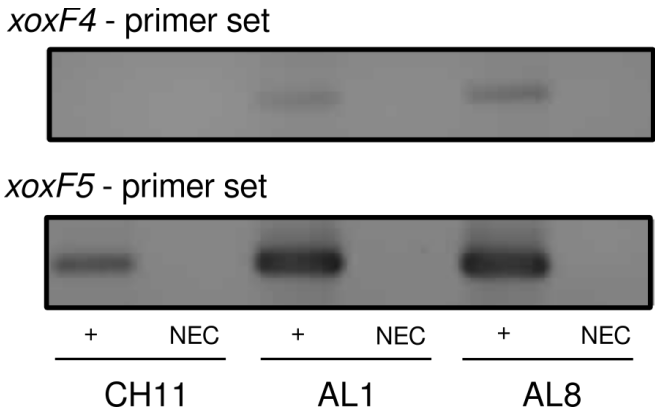

FiG S4.

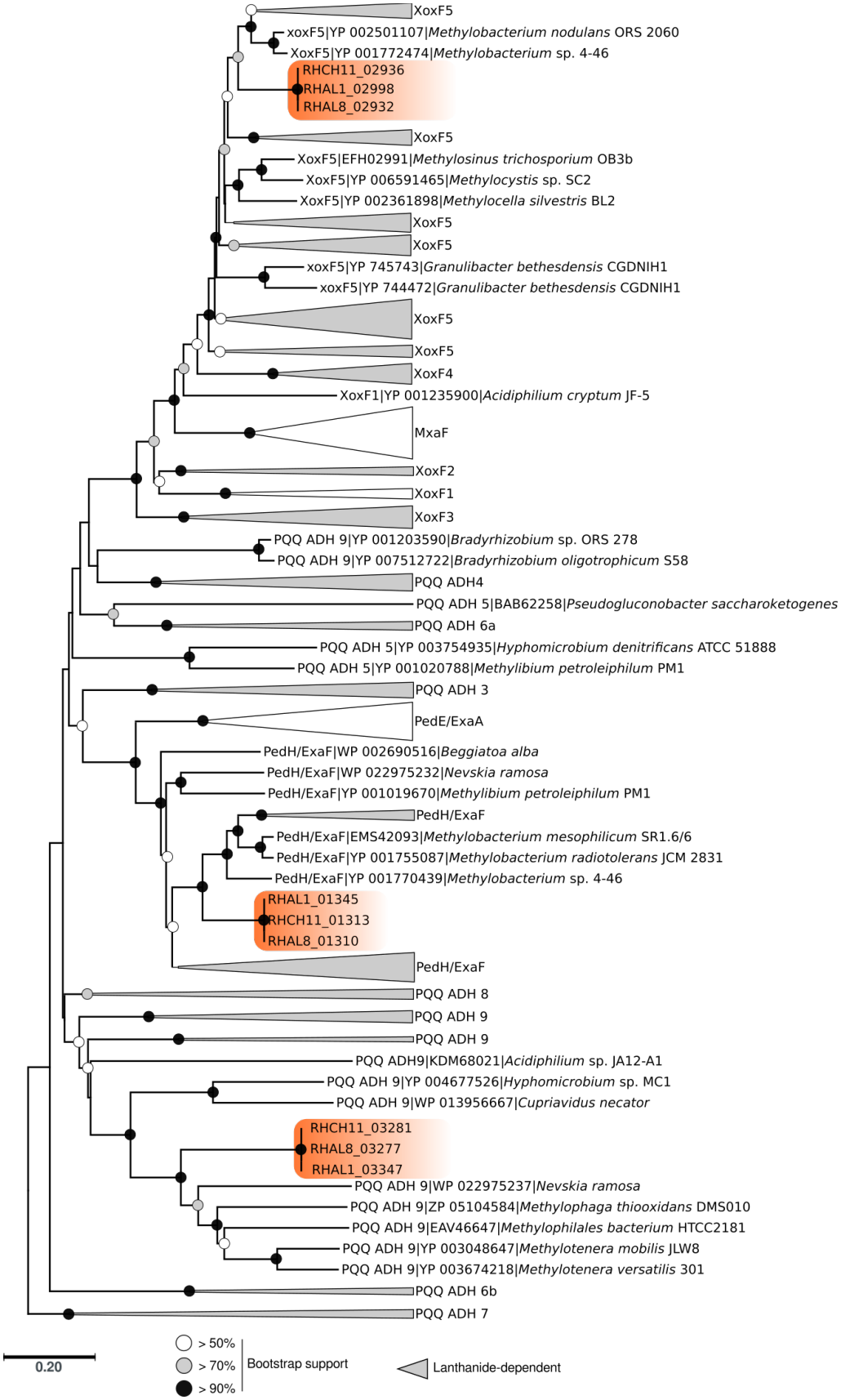
XoxF5
xoxF5|YP 002501107|Methylobacterium nodulans ORS 2060
XoxF5|YP 001772474|Methylobacterium sp. 4-46
RHCH11_02936
RHAL1_02998
RHAL8_02932
XoxF5
XoxF5|EFH02991|Methylosinus trichosporium OB3b
XoxF5|YP 006591465|Methylocystis sp. SC2
XoxF5|YP 002361898|Methylocella silvestris BL2
XoxF5
XoxF5
xoxF5|YP 745743|Granulibacter bethesdensis CGDNIH1
xoxF5|YP 744472|Granulibacter bethesdensis CGDNIH1
XoxF5
XoxF5
XoxF4
XoxF1|YP 001235900|Acidiphilium cryptum JF-5
MxaF
XoxF2
XoxF1
XoxF3
PQQ ADH 9|YP 001203590|Bradyrhizobium sp. ORS 278
PQQ ADH 9|YP 007512722|Bradyrhizobium oligotrophicum S58
PQQ ADH4
PQQ ADH 5|BAB62258|Pseudogluconobacter saccharoketogenes
PQQ ADH 6a
PQQ ADH 5|YP 003754935|Hyphomicrobium denitrificans ATCC 51888
PQQ ADH 5|YP 001020788|Methylibium petroleiphilum PM1
PQQ ADH 3
PedE/ExaA
PedH/ExaF|WP 002690516|Beggiatoa alba
PedH/ExaF|WP 022975232|Nevskia ramosa
PedH/ExaF|YP 001019670|Methylibium petroleiphilum PM1
PedH/ExaF
PedH/ExaF|EMS42093|Methylobacterium mesophilicum SR1.6/6
PedH/ExaF|YP 001755087|Methylobacterium radiotolerans JCM 2831
PedH/ExaF|YP 001770439|Methylobacterium sp. 4-46
RHAL1_01345
RHCH11_01313
RHAL8_01310
PedH/ExaF
PQQ ADH 8
PQQ ADH 9
PQQ ADH 9
PQQ ADH9|KDM68021|Acidiphilium sp. JA12-A1
PQQ ADH 9|YP 004677526|Hyphomicrobium sp. MC1
PQQ ADH 9|WP 013956667|Cupriavidus necator
RHCH11_03281
RHAL8_03277
RHAL1_03347
PQQ ADH 9|WP 022975237|Nevskia ramosa
PQQ ADH 9|ZP 05104584|Methylophaga thiooxidans DMS010
PQQ ADH 9|EAV46647|Methylophilales bacterium HTCC2181
PQQ ADH 9|YP 003048647|Methylotenera mobilis JLW8
PQQ ADH 9|YP 003674218|Methylotenera versatilis 301
PQQ ADH 6b
PQQ ADH 7
0.20
> 50%
> 70%
> 90%
Bootstrap support
Lanthanide-dependent

Fig S5.

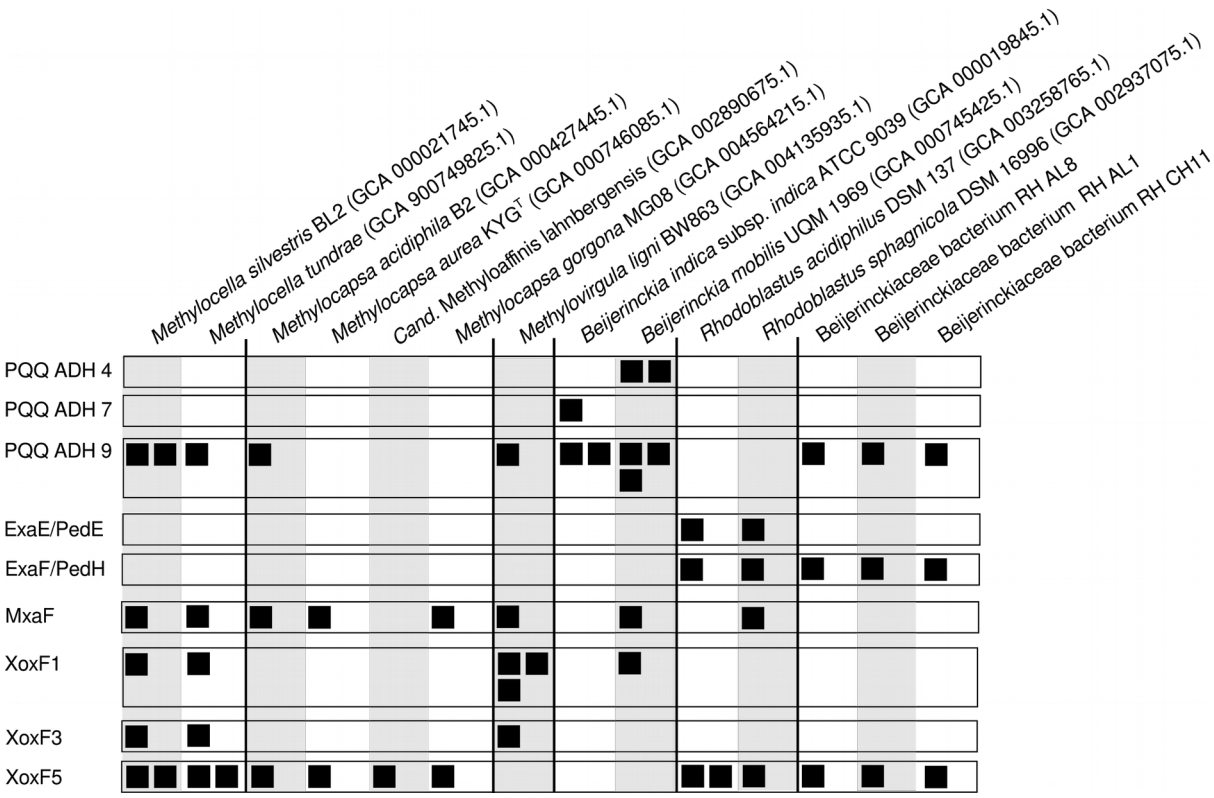

Fig S6.

Calmodulin (*Homo sapiens*, NP_005175.2)

```
EF1 21   DKDGDGTITTKE 32
EF2 57   DADGNGTIDFPE 68
EF3 94   DKDGNGYISAAE 105
EF4 130  DIDGDGQVNYEE 141
```

Lanmodulin (*Methylorubrum extorquens* AM1, META1_1786)

```
EF1 34   DPDKDGTIDLKE 45
EF2 58   DPDKDGTLDAKE 69
EF3 83   DPDNDGTLDKKE 95
EF4 107  NPDNDGTUDARE 118
```

Lanmodulin-homolog
(Beijerinckiaceae bacterium RH AL1, RHAL1_01396)

```
EF1 35   DPDSDGTVSLAE 46
EF2 59   DPDNDGTIDLKE 70
EF3 84   DADNDGTVDKAE 95
EF4 108  DPDGDGTLDAKE 119
```

Fig S7.

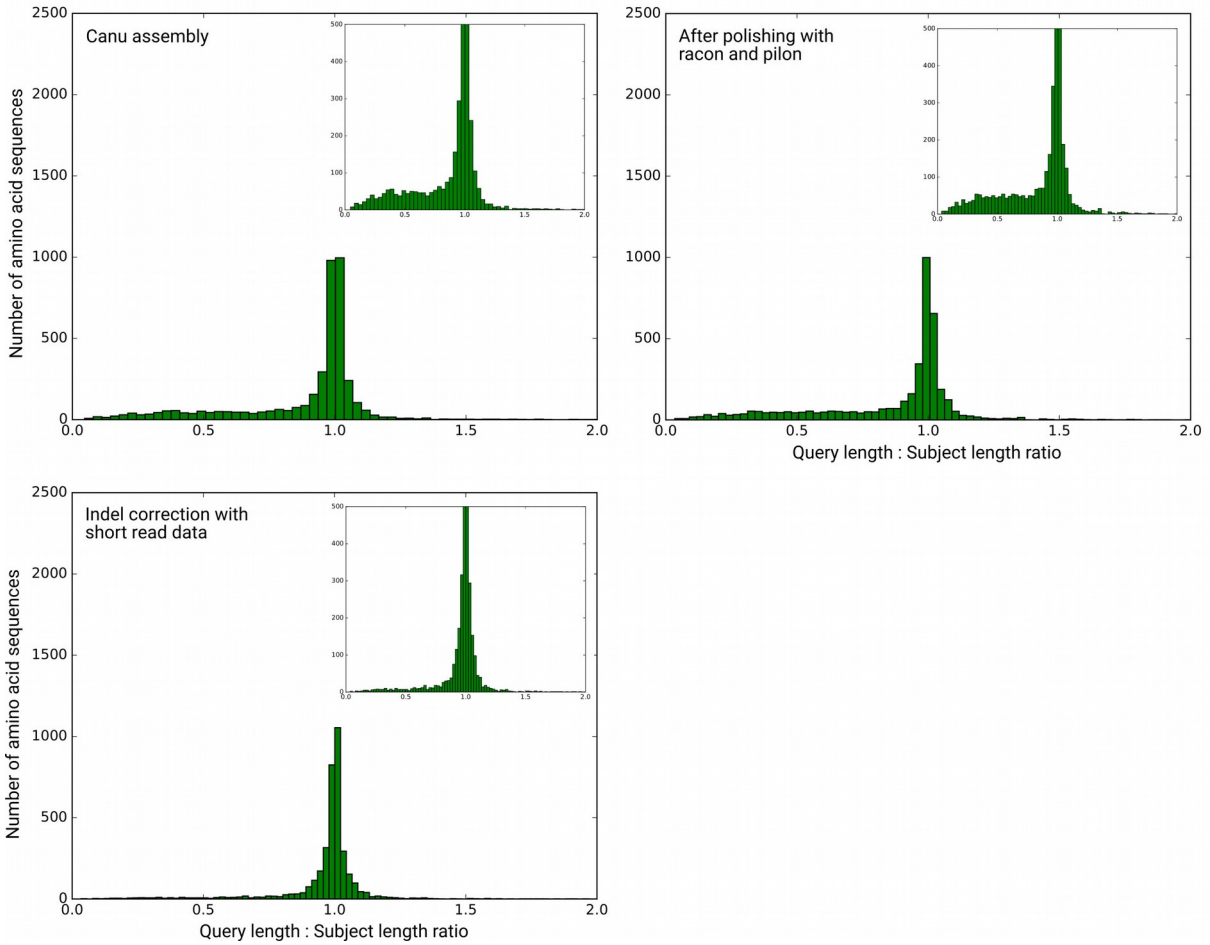
Canu assembly
After polishing with racon and pilon
Indel correction with short read data
Number of amino acid sequences
Query length : Subject length ratio

Fig S8.

REFERENCES

1. Qin QL, Xie BB, Zhang XY, Chen XL, Zhou BC, Zhou J, Oren A, Zhang YZ. 2014. A proposed genus boundary for the prokaryotes based on genomic insights. J. Bacteriol. 196:2210-2215.

2. Taubert M, Grob C, Howat AM, Burns OJ, Dixon JL, Chen Y, Murrell JC. 2015. XoxF encoding an alternative methanol dehydrogenase is widespread in coastal marine environments. Environ. Microbiol. 17:3937-3948.

3. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

4. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. 35:1547–1549.

5. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

6. Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.

7. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27:722-736.

8. Buchfink , Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12:59–60.

9. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27:737–746.

10. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. Pilon: an integrated tool for comprehensive

microbial variant detection and genome assembly improvement. PLoS One. 9:e112963.

11. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics, 28:1823–1829.

12. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41:D590–596.

13. Zeien H, Brümmer GW. 1989. Chemische Extraktion zur Bestimmung der Schwermetallbindungsformen in Böden. Mittlgn. Dtsch. Bodenkundl. Gesellsch. Sonderh. ; 59/I, Kongreßbd. Münster:505-510.

14. Weber J, Dradrach A, Karczewska A, Kocowicz A. 2018. The distribution of sequentially extracted Cu, Pb, and Zn fractions in Podzol profiles under dwarf pine of different stages of degradation in subalpine zone of Karkonosze Mts (central Europe). J. Soils Sediments 18:2387-2398.

15. Wegner CE, Liesack W. 2017. Unexpected Dominance of Elusive Acidobacteria in Early Industrial Soft Coal Slags. Front. Microbiol. 8:1013.

16. Keltjens JT, Pol A, Reimann J, Op Den Camp HJM. 2014. PQQ-dependent methanol dehydrogenases: Rare-earth elements make a difference. Appl. Microbiol. Biotechnol. 98:6163-6183.

17. Eddy SR. 2011. Accelerated Profile HMM Searches. PLoS Comput. Biol. 7:e1002195.

18. Sait M, Hugenholtz P, Janssen PH. 2002. Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. Environ. Microbiol. 4:654-666.

19. Dedysh SN, Dunfield PF. 2014. Cultivation of methanotrophs. In: McGenity T, Timmis K, Nogales B, eds. Hydrocarbon and lipid microbiology protocols, Springer protocols handbooks. Berlin: Springer-Verlag.

20. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

21. Bourne DG, McDonald IR, Murrell JC. 2001. Comparison of *pmoA* PCR primer sets as tools for investigating methanotroph diversity in three danish soils. Appl. Environ. Microbiol. 67:3802-3809.

22. Holmes AJ, Costello A, Lidstrom ME, Murrell JC. 1995. Evidence that particulate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. FEMS Microbiol. Lett. 132:203–208.

23. Costello AM, Lidstrom ME. 1999. Molecular characterization of functional and phylogenetic genes from natural populations of methanotrophs in lake sediments. Appl. Environ. Microbiol. 65:5066–5074.

24. Oshkin IY, Wegner CE, Lüke C, Glagolev MV, Filippov IV, Pimenov NV, Liesack W, Dedysh SN. 2014. Gammaproteobacterial methanotrophs dominatecold methane seeps in floodplains of west siberian rivers. Appl. Environ. Microbiol. 80:5944–5954.

25. Dunfield PF, Belova SE, Vorob'ev AV, Cornish SL, Dedysh SN. 2010. *Methylocapsa aurea* sp. nov., a facultative methanotroph possessing a particulate methane monooxygenase, and emended description of the genus *Methylocapsa*. Int. J. Syst. Evol. Microbiol. 60:2659-2664.

26. McDonald IR, Murrell JC. 1997. The methanol dehydrogenase structural gene *mxaF* and its use as a functional gene probe for methanotrophs and methylotrophs. Appl. Environ. Microbiol. 63:3218–3224.

27. Lau E, Fisher MC, Steudler PA, Cavanaugh CM. 2013. The methanol dehydrogenase gene, *mxaF*, as a functional and phylogenetic marker for proteobacterial methanotrophs in natural environments. PLoS One, 8:e56993.

28. Neufeld JD, Schäfer H, Cox MJ, Boden R, McDonald IR, Murrell JC. 2007. Stable-isotope probing implicates Methylophaga spp and novel Gammaproteobacteria in marine methanol and methylamine metabolism. ISME J. 1:480-491.

29. Wischer D, Kumaresan D, Johnston A, Khawand ME, Stephenson J, Hillebrand-Voiculescu AM, Chen Y, Murrell JC. Bacterial metabolism of methylated amines and identification of novel methylotrophs in movile cave. ISME J. 9:1–12.

30. Antony CP, Kumaresan D, Ferrando L, Boden R, Moussard H, Scavino AF, Shouche YS, Murrell JC. 2010. Active methylotrophs in the sediments of lonar lake, a saline and alkaline ecosystem formed by meteor impact. ISME J. 4:1470–1480.

31. Venable JH, Coggeshall R. 1965. A simplified lead citrate stain for use in electron microscopy. J. Cell Biol., 25:407–408.

32. Bushnell B. 2016. BBMap short read aligner.2016. https://www.sourceforge.net/projects/bbmap/.

33. ABankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev Ma, Pevzner PA. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. 19:455–477.

34. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5:R12.

35. Hunt M, De Silva N, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol. 16:294.

36. Kolmogorov M, Raney B, Paten B, Pham S. 2014. Ragout-a reference-assisted assembly tool for bacterial genomes. Bioinformatics 30:I302–I309.

37. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319.

38. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD.2019. The pfam protein families database in 2019. Nucleic Acids Res. 47:D427–D432.

39. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28:33–36.

40. Benedict MN, Henriksen JR, Metcalf WW, Whitaker RJ, Price ND. 2014. ITEP: An integrated toolkit for exploration of microbial pan-genomes. BMC Genomics, 15: 8.

41. van Dongen S, Abreu-Goodger C. 2012. Using MCL to extract clusters from networks. Methods Mol. Biol. 804:281–295.

42. Felsenstein J. 1989. PHYLIP-Phylogeny inference package (version 3.2). Cladistics, 5:164–166.

43. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

44. Orata FD, Meier-Kolthoff JP, Sauvageau D, Stein LY. 2018. Phylogenomic analysis of the gammaproteobacterial methanotrophs (order methylococcales) calls for the reclassification of members at the genus and species levels. Front. Microbiol., 9:3162.