Supplementary materials

Metagenomic insights into ecological and phylogenetic significances of *Candidatus* Natranaeroarchaeales, a novel abundant archaeal order in soda lake sediment

Heng Zhou[1†], Dahe Zhao[1†*], Shengjie Zhang[1,2], Qiong Xue[1,2], Manqi Zhang[1,2], Haiying Yu[1], Jian Zhou[1], Ming Li[1], Hua Xiang[1,2*]

[1] State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China

[2] College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China

* Correspondance: Hua Xiang, xiangh@im.ac.cn; Dahe Zhao, zhaodh@im.ac.cn

[†] These authors have contributed equally to this work.

Running title: Novel abundant archaeal order in soda lake sediment

Supplementary Results

Putative transporters

Ca. Natraneroarchaeales contained a quantity of transport systems (Fig. S7). Most MAGs (16 out of 25 MAGs) contained trk system potassium uptake protein (TrkAHG), which was essential for the osmotic pressure balance. Six high-quality and two medium-quality MAGs had all seven genes of Multicomponent Na^+: H^+ antiporter (*mnhABCDEFG*), while ten high- and medium-MAGs had six genes (lacking *mnhA*). Mnh is of importance in the intracellular pH homeostasis. Glycine betaine/proline transport system (ProVWX) was only found in MAG B1Sed10_20. About half of the MAGs (including 8 high-quality and 3 medium-quality MAGs) contained phosphate transport system (PstABCS) which was especially important for microbes where phosphorus was scarce. Zinc transport system (AnuABC) was found in ten high- and medium-quality MAGs. Molybdate/tungstate transport system (WtpABC) was only found in T1Sed10_119m and B1Sed_34, which was presented mainly in enzymes that catalyze oxygen atom transfer reactions. Teichoic acid transport system permease protein (TagGH) was only found in CSSed10_214.

Metabolic potentials of carbohydrate, nucleotide, amino acid and membrane lipid

Almost all high-quality MAGs contained most genes of Embden-Meyerhof pathway (from glucose-6-phosphate to pyruvate) except 6-phosphofructokinase gene (*pfk*), which blocked the carbohydrate degradation (Fig. 5 & S9). Meanwhile, pyruvate/orthophosphate dikinase (*ppdK*), pyruvate water dikinase (*ppsA*) and fructose 1,6-bisphosphate aldolase/phosphatase (*fbp*) were present in most high-quality MAGs (Fig. 5 & S9). It indicated that carbon flux may flow via the gluconeogenesis direction. The non-oxidative (transketolase and transaldolase) and oxidative

(3-hexulose-6-phosphate synthase, 6-phospho-3-hexuloisomerase, ribose 5-phosphate isomerase and ribulose-phosphate 3-epimerase) processes of phosphate pentose pathway were also present in most high-quality MAGs (Fig. 5 & S9). It suggested that phosphate pentose pathway may function in carbohydrate utilization. Ribose-phosphate pyrophosphokinase (PrsA) could catalyze the formation of 5-phosphoribosyl 1-pyrophosphate, which was the precursor of nucleotide biosynthesis (Fig. 5). Six high-quality MAGs (CSSed10_214, HAT26, CSSed10_245, B1Sed10_107R1, B1Sed10_20 and T1Sed10_8R1) contained all genes included in the *de novo* biosynthesis of purine, while high-quality MAG HAT25 have complete pyrimidine biosynthesis pathway (Fig. S10). Besides, 5 MAGs lacked only one or two genes (*purD*, *purC*, *purP* or *purB*) for purine biosynthesis.

The amino acid metabolism in MAGs with high and medium quality was showed in Fig. S11. Most high-quality (ten out of eleven) and one medium quality MAGS contained glutamate dehydrogenase (*gudB*, *gdhA* and *gdhA2*), glutamate synthase (NADPH) small chain (*gltD*) and glutamine synthetase (*glnA*) who function in ammonia assimilation and metabolism of glutamate and glutamine. Most of high- and medium-quality MAGs (23 out of 25 MAGs) contained aspartate aminotransferase (*aspB*), which catalyzed the transformation between oxaloacetate and aspartate. Asparagine synthase (glutamine-hydrolysing) (*asnB*) was the key gene in the formation of asparagine which was found in 3 high- and 5 medium-quality MAGs. Most MAGs contained gene asparaginase (*ansAB*) by which asparagine could be transformed into aspartate. Besides, most high-quality MAGs (ten out of eleven) and half medium-quality MAGs encoded serine O-acetyltransferase (*cysE*) and cysteine synthase (*cysK*), which catalyzed the biosynthesis of cysteine from serine (sulfur assimilation). Almost all MAGs (except three medium-quality MAGs)

had threonine aldolase gene (*ltaE*) which participated in the biosynthesis of glycine from threonine.

Mevalonate pathway for terpenoid backbone biosynthesis was almost complete in *Ca.* Natranaeroarchaeales. Presence of mevalonate kinase (MVK in most MAGs) indicated that mevalonate 5-phosphate was the intermediate. There were two pathways for the biosynthesis of isopentenyl-diphosphate from mevalonate 5-phosphate, via isopentenyl-phosphate or mevalonate 5-diphosphate (Fig. S12a). Most MAGs contained the gene for isopentenyl phosphate kinase (*ipk*), and it suggested the biosynthesis via isopentenyl-phosphate (Fig. S12b). Another hypothetical protein may function similarly as phosphomevalonate decarboxylase (Pmd).

Thio: fumarate reductase / fumarate reductase (CoM/CoB)

There were twelve complete and one truncated (at one terminal of contig in HAT29) *tfrA* genes found in *Ca.* Natranaeroarchaeales (Fig. S13). Two *tfrA* fragment were found in the terminal of contigs PWEC01000073.1 and PWEC01000107.1 in MAG T1Sed10_92 (Table S9), and both contigs shared 120 bp nucleotide sequence at *tfrA* gene. So, they were merged in the figure. Six of these genes formed an operon with *tfrB*. Considering the genes (*tfrB* and unknown function genes by eggNOG) between *hdrB* and *tfrA* were similar length, and they shared high identities (more than 70%, shown in Fig. S14), so these function unknown genes were named *tfrB*-homologs.

Abbreviation

Enzyme definition:

PurF: amidophosphoribosyltransferase [EC:2.4.2.14]
PurD: phosphoribosylamine---glycine ligase [EC:6.3.4.13]
PurCD: fusion protein PurCD [EC:6.3.2.6 6.3.4.13]
PurN: phosphoribosylglycinamide formyltransferase 1 [EC:2.1.2.2]
Pfas: phosphoribosylformylglycinamidine synthase [EC:6.3.5.3]
PurM: phosphoribosylformylglycinamidine cyclo-ligase [EC:6.3.3.1]
PurE: 5-(carboxyamino)imidazole ribonucleotide mutase [EC:5.4.99.18]
PurB: adenylosuccinate lyase [EC:4.3.2.2]
PurH: phosphoribosylaminoimidazolecarboxamide formyltransferase / IMP cyclohydrolase [EC:2.1.2.3 3.5.4.10]
PurA: adenylosuccinate synthase [EC:6.3.4.4]
CarAB: carbamoyl-phosphate synthase small/large subunit [EC:6.3.5.5]
PyrBI: aspartate carbamoyltransferase catalytic/regulatory subunit [EC:2.1.3.2]
Ura: dihydroorotase [EC:3.5.2.3]
PyrD: dihydroorotate dehydrogenase (fumarate) [EC:1.3.98.1]
Dhodh: dihydroorotate dehydrogenase [EC:1.3.5.2]
PyrDI: dihydroorotate dehydrogenase (NAD^+) catalytic subunit [EC:1.3.1.14]
Umps: uridine monophosphate synthetase [EC:2.4.2.10 4.1.1.23]
PyrFE: orotidine-5'-phosphate decarboxylase [EC:4.1.1.23]/orotate phosphoribosyltransferase [EC:2.4.2.10]
PyrG: CTP synthase [EC:6.3.4.2]
PyrH: uridylate kinase [EC:2.7.4.22]
Adk, Ak: adenylate kinase [EC:2.7.4.3]
Ndk: nucleoside-diphosphate kinase [EC:2.7.4.6]
PurK: 5-(carboxyamino)imidazole ribonucleotide synthase [EC:6.3.4.18]
PAICS: phosphoribosylaminoimidazole carboxylase / phosphoribosylaminoimidazole-succinocarboxamide synthase [EC:4.1.1.21 6.3.2.6]
Ade: phosphoribosylaminoimidazole carboxylase [EC:4.1.1.21]
Pk: pyruvate kinase [EC:2.7.1.40]
ACAT: acetyl-CoA C-acetyltransferase [EC:2.3.1.9]
Hmcs: hydroxymethylglutaryl-CoA synthase [EC:2.3.3.10]
MvaA: hydroxymethylglutaryl-CoA reductase [EC:1.1.1.88]
Mvk: mevalonate kinase [EC:2.7.1.36]
Ipk: isopentenyl phosphate kinase [EC:2.7.4.26]
Idi: isopentenyl-diphosphate Delta-isomerase [EC:5.3.3.2]
IdsA: geranylgeranyl diphosphate synthase, type I [EC:2.5.1.1 2.5.1.10 2.5.1.29]
Mvd: diphosphomevalonate decarboxylase [EC:4.1.1.33]
Pmd: phosphomevalonate decarboxylase [EC:4.1.1.99]
PmvK: phosphomevalonate kinase [EC:2.7.4.2]
Rnf: Na+-translocating ferredoxin:NAD^+ oxidoreductase subunit ABCDEG

ATP: V/A-type H^+/Na^+-transporting ATPase subunit ABCDEFIK [EC:7.1.2.2 7.2]
PhaFG: multicomponent K^+:H^+ antiporter subunit FG
MnhABCDEFG: multicomponent Na^+:H^+ antiporter subunit ABCDEFDG
Nuo: NADH-quinone oxidoreductase subunit EFIKLMN [EC:7.1.1.2]
NirAB: ferredoxin-nitrite reductase [EC:1.7.7.1]/nitrite reductase (NADH) large subunit [EC:1.7.1.15]
PAPSS: 3'-phosphoadenosine 5'-phosphosulfate synthase [EC:2.7.7.4 2.7.1.25]
Sat: sulfate adenylyltransferase [EC:2.7.7.4]
CysC: adenylylsulfate kinase [EC:2.7.1.25]
CysH: phosphoadenosine phosphosulfate reductase [EC:1.8.4.8 1.8.4.10]
Sir: sulfite reductase (ferredoxin) [EC:1.8.7.1]
HydGBAD: sulfhydrogenase subunit gamma/beta(sulfur reductase)/alpha/delta [EC:1.12.98.4]
PhsA: thiosulfate reductase / polysulfide reductase chain A [EC:1.8.5.5]
TST: thiosulfate/3-mercaptopyruvate sulfurtransferase [EC:2.8.1.1 2.2]
CooS, AcsA: anaerobic carbon-monoxide dehydrogenase catalytic subunit [EC:1.2.7.4]
FdhAB: formate dehydrogenase ($NADP^+$) alpha/beta subunit [EC:1.17.1.10]
Fhs: formate--tetrahydrofolate ligase [EC:6.3.4.3]
FolD: methylenetetrahydrofolate dehydrogenase ($NADP^+$) / methenyltetrahydrofolatecyclohydrolase [EC:1.5.1.5 3.5.4.9]
MetF: methylenetetrahydrofolate reductase (NADPH) [EC:1.5.1.20]
CdhDE: acetyl-CoA decarbonylase/synthase, CODH/ACS complex subunit delta/gamma [EC:2.1.1.245]
AcsB: acetyl-CoA synthase [EC:2.3.1.169]
Ppc: phosphoenolpyruvate carboxylase [EC:4.1.1.31]
Oad: oxaloacetate decarboxylase (Na^+ extruding) subunit alpha [EC:7.2.4.2]
PycAB: pyruvate carboxylase subunit A/B [EC:6.4.1.1]
Me: malate dehydrogenase (oxaloacetate-decarboxylating) [EC:1.1.1.38]
MaeB: malate dehydrogenase (oxaloacetate-decarboxylating)($NADP^+$) [EC:1.1.1.40]
FumAB: fumarate hydratase subunit alpha/beta [EC:4.2.1.2]
TfrAB: fumarate reductase (CoM/CoB) subunit AB [EC:1.3.4.1]
SucCD: succinyl-CoA synthetase [EC:6.2.1.5]
KorDABC: 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase subunit delta/alpha/beta/gamma [EC:1.2.7.3 1.2.7.11]
Idh: isocitrate dehydrogenase [EC:1.1.1.42]
Aco: aconitate hydratase [EC:4.2.1.3]
AclAB: ATP-citrate lyase alpha/beta-subunit [EC:2.3.3.8]
Mdh: malate dehydrogenase [EC:1.1.1.37]
Mvh: F420-non-reducing hydrogenase large/small /iron-sulfur subunit [EC:1.12.99.-1.8.]
Hdr: heterodisulfide reductase subunit ABC [EC:1.8.7.3 1.8.98.4 1.8.98.5 1.8.98]
AcyP: acylphosphatase [EC:3.6.1.7]
AcdAB: acetate---CoA ligase (ADP-forming) subunit alpha/beta [EC:6.2.1.13]
Acs: acetyl-CoA synthetase [EC:6.2.1.1]
Pta: phosphate acetyltransferase [EC:2.3.1.8]
Por: pyruvate ferredoxin oxidoreductase alpha/beta/delta/gamma subunit [EC:1.2.7.1]

Pgk: phosphoglycerate kinase [EC:2.7.2.3]
Gpm: 2,3-bisphosphoglycerate-independent phosphoglycerate mutase [EC:5.4.2.12]
Eno: enolase [EC:4.2.1.11]
Pyk: pyruvate kinase [EC:2.7.1.40]
PpdK: pyruvate, orthophosphate dikinase [EC:2.7.9.1]
PpsA: pyruvate, water dikinase [EC:2.7.9.2]
PoxL: pyruvate oxidase [EC:1.2.3.3]
Fba: fructose-bisphosphate aldolase, class I [EC:4.1.2.13]
Pgi: glucose-6-phosphate isomerase, archaeal [EC:5.3.1.9]
Tpi: triosephosphate isomerase (TIM) [EC:5.3.1.1]
Gap: glyceraldehyde-3-phosphate dehydrogenase (NAD(P)) [EC:1.2.1.59]
TktA: transketolase [EC: 2.2.1.1]
TalA: transaldolase [EC: 2.2.1.2]
Rpe: ribulose-phosphate 3-epimerase [EC:5.1.3.1]
RpiA: ribose 5-phosphate isomerase A [EC:5.3.1.6]
HxlBA: 6-phospho-3-hexuloisomerase [EC:5.3.1.27]/3-hexulose-6-phosphate synthase [EC:4.1.2.43]
Hps-Phi: 3-hexulose-6-phosphate synthase / 6-phospho-3-hexuloisomerase [EC:4.1.2.1.27]
PrsA: ribose-phosphate pyrophosphokinase [EC:2.7.6.1]
Pfk: ATP-dependent phosphofructokinase / diphosphate-dependent phosphofructokinase [EC:2.7.1.11 2.7.1.90]
Fbp: fructose 1,6-bisphosphate aldolase/phosphatase [EC:4.1.2.13 3.1.3.11]
ZnuABC: zinc transport system
PstABCS: phosphate transport system
TrkHGA: trk system potassium uptake protein

Compounds abbreviation:
Glucose-6P: Glucose 6-phosphate
Fructose-6P: Fructose 6-phosphate
Fructose-1,6P_2: Fructose 1,6-bisphosphate
Glyceraldehyde-3P: Glyceraldehyde 3-phosphate
Glycerate-1,3P_2: 3-Phospho-D-glyceroyl phosphate
Glycerone-P: Glycerone phosphate
Ribulose-5P: Ribulose 5-phosphate
Ribose-5P: Ribose 5-phosphate
Xylulose-5P: Xylulose 5-phosphate
PRPP: 5-Phosphoribosyl 1-pyrophosphate
IMP: Inosine monophosphate
ADP: Adenosine 5'-diphosphate
ATP: Adenosine 5'-triphosphate
UMP: Uridine monophosphate
UDP: Uridine 5'-diphosphate
UTP: Uridine triphosphate
CDP: Cytidine diphosphate

CTP: Cytidine triphosphate
PEP: Phosphoenolpyruvate
IPP: Isopentenyl diphosphate
GGPP: Geranylgeranyl diphosphate
THF: Tetrahydrofolate
APS: Adenylyl sulfate
PAPS: 3'-Phosphoadenylyl sulfate
NAD^+: Nicotinamide adenine dinucleotide
NADH: Reduced nicotinamide adenine dinucleotide.
CPRI: 5-Carboxyamino-1-(5-phospho-D-ribosyl)imidazole
GAR: 5'-Phosphoribosylglycinamide
FGAR: 2-(Formamido)-N1-(5'-phosphoribosyl)acetamidine
AIR: Aminoimidazole ribotide
CAIR: 1-(5-Phospho-D-ribosyl)-5-amino-4-imidazolecarboxylate
SAICAR: 1-(5'-Phosphoribosyl)-5-amino-4-(N-succinocarboxamide)-imidazole
AICAR: 1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide
FAICAI: 1-(5'-Phosphoribosyl)-5-formamido-4-imidazolecarboxamide
AMP: Adenosine 5'-monophosphate

Supplementary figures

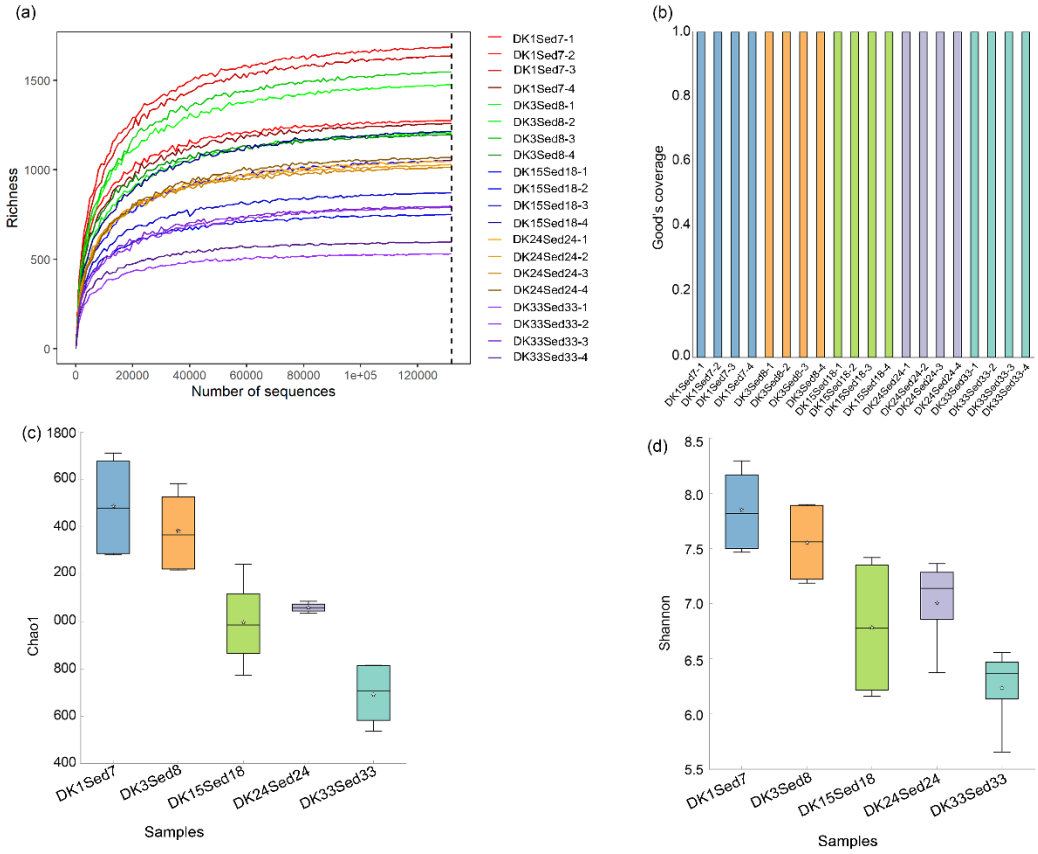

Fig. S1 Alpha diversity based on 16S rRNA amplicon in the deep sediment of five soda lake ponds. The archaea-specific primers (349F and 806R) were used to obtain the V3-V4 region of 16S rRNA gene. Richness (a) and Good's coverage (b) indicated the sequencing depth of twenty samples was enough. Among the twenty samples, DK and Sed indicated the Habor Lake and sediment samples. The numbers after DK and Sed showed the salinity of brine and that of pore water in the sediment. Four duplicates were designed for each sediment. Chao1 (c) and Shannon (d) indexes exhibited the alpha diversity of the five sediment.

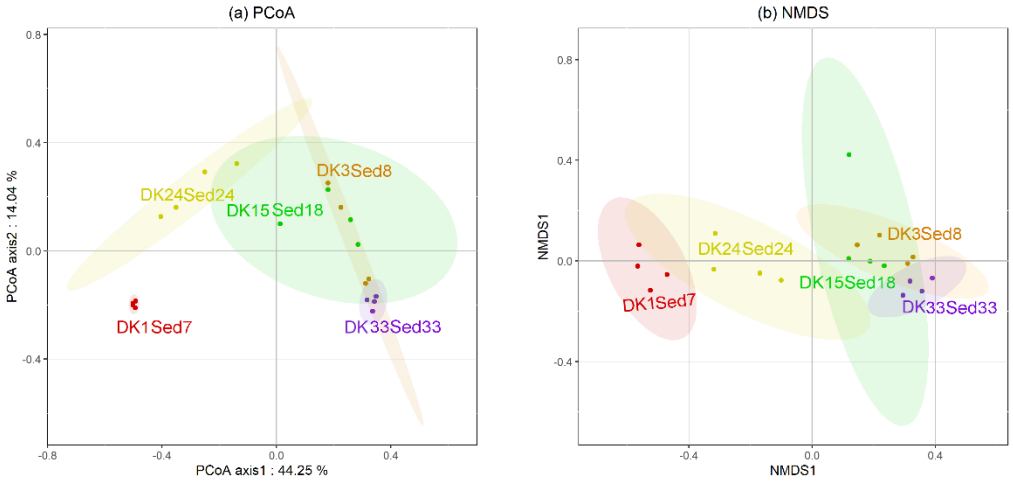

Fig. S2 Beta diversity based on 16S rRNA amplicon in the deep sediment of five soda lake ponds. The principal coordinate analysis (PCoA) (a) and Non-metric multidimensional scaling (NMDS) plots (b) were performed based on the Bray-Curtis distances. The confidence ellipse was 95%.

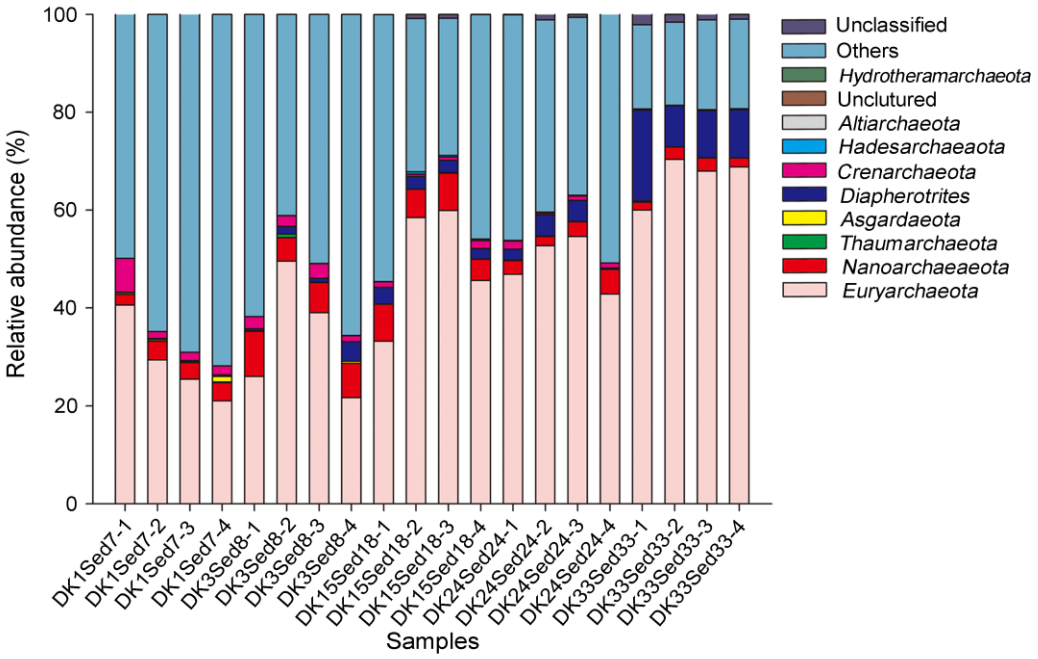

Fig. S3 Archaea community composition profiles of deep sediment in soda lake with different salinity. The top lineages at phylum levels were exhibited based on the SILVA database release 136.

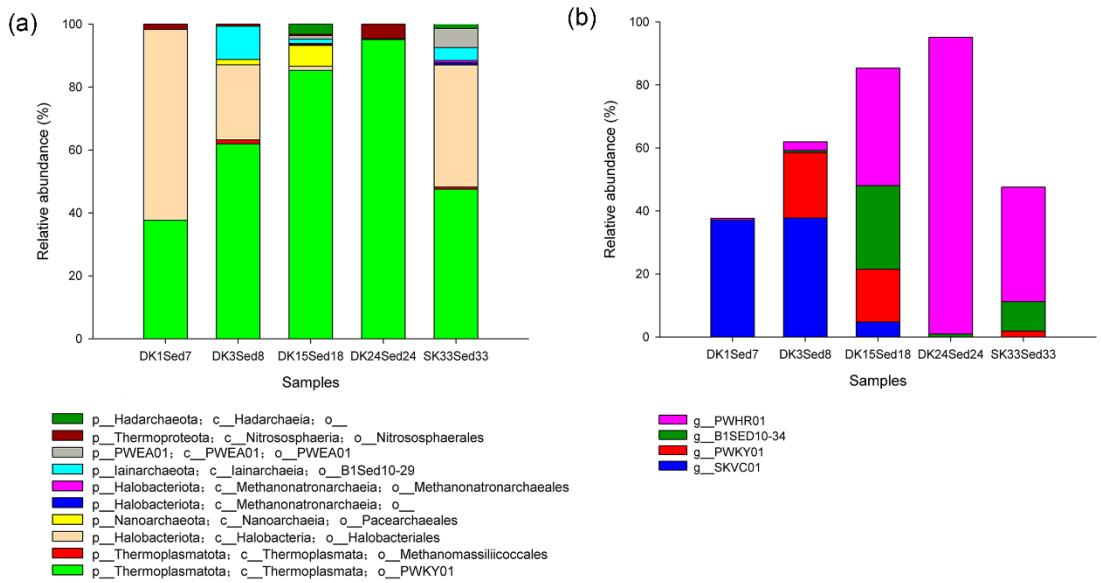

Fig. S4 Microbial community compositions based on the relative abundance of metagenome-assembled genomes in the deep sediment of five soda lake ponds. (a) The relative abundance of archaea was merged at order level. (b) The relative abundance of phylum *Thermoplasmatota* (in GTDB) was merged at genus level.

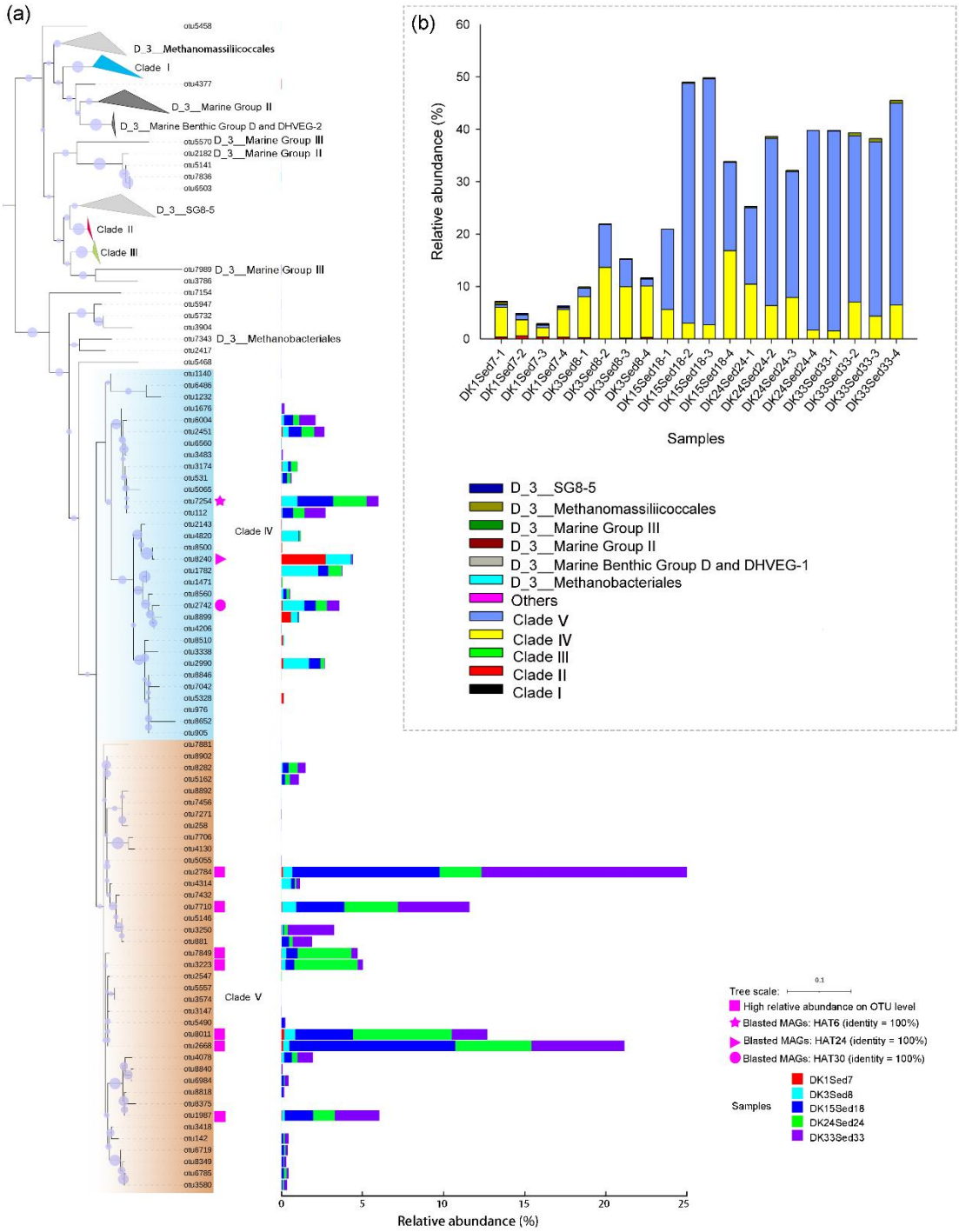

Fig. S5 Phylogenetic tree and relative abundance of class *Thermoplasmata* (in SILVA). (a) Maximum likelihood phylogeny with 1000 times bootstrap was constructed based on the 16s rRNA amplicon sequences. The unclassified *Thermoplasmata* sequences was separated into clades I to V. Tree scale, 0.1; pink square, OTU of high abundance; pink star, triangular and circle, OTU shared a 100% identity with the 16S rRNA gene from MAGs (data shown in Table S6). (b) The relative abundance of the five *Thermoplasmata* clades based on the feature count of 16s rRNA amplicon sequences.

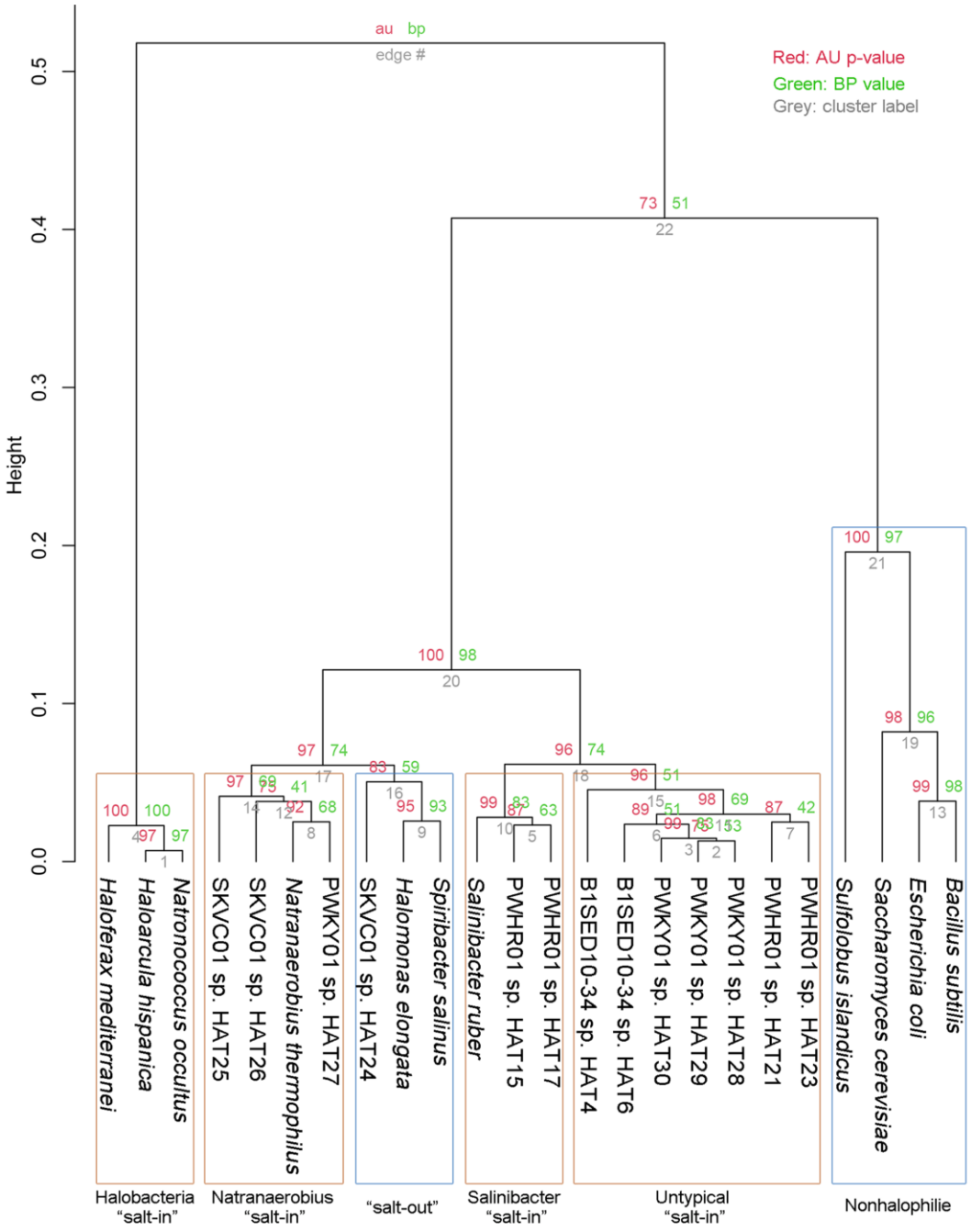

Fig. S6 Cluster analysis of the isoelectric point profiles of predicted proteomes. The *Ca.* Natranaeroarchaeales MAGs of high- (completeness > 90%, contamination < 5%) and medium-quality (completeness > 70%, contamination < 10%) were included. Red and green number indicated the approximately unbiased p-value and bootstrap probability value (both are percentages), respectively. Bootstrap resampling was 1000 times. Grey number labeled the cluster. Isoelectric point (pI) profiles were cluster into six classes: I, salt-out; II, Natranaerobius salt-in; III, Untypical salt-in; IV, Salinibacter salt-in; V, Halobacteria salt-in; VI, Nonhalophile.

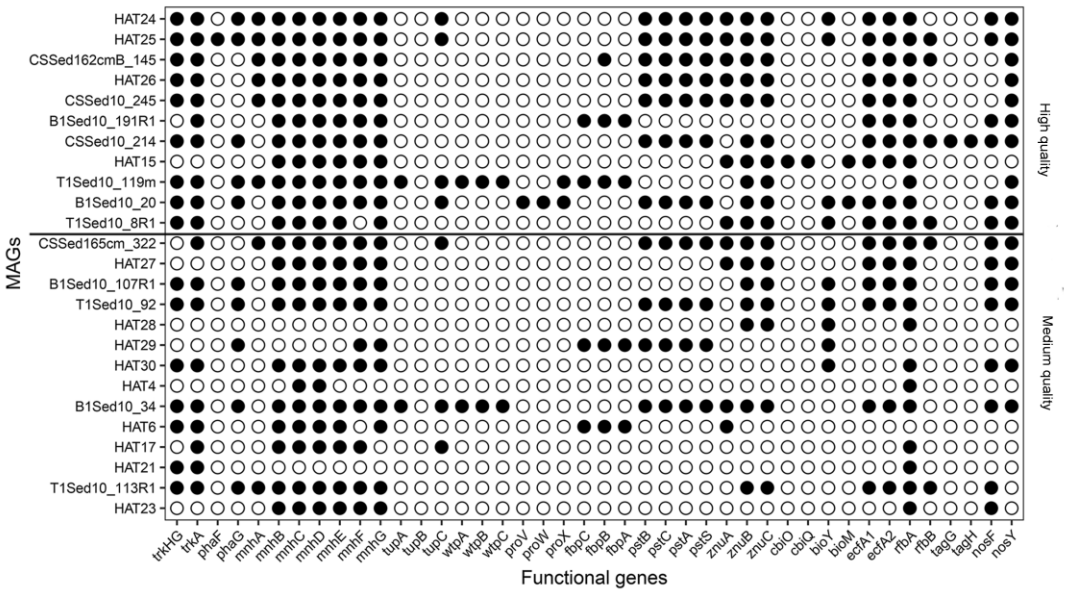

Fig. S7 Presence of genes involved in substrate transporters in *Ca.* Natranaeroarchaeales MAGs. The high- (completeness > 90%, contamination < 5%) and medium- quality (completeness > 70%, contamination < 10%) MAGs were included. Solid and hollow dots indicated the presence and absence in the MAGs, respectively. The definitions of genes were listed in Table S8.

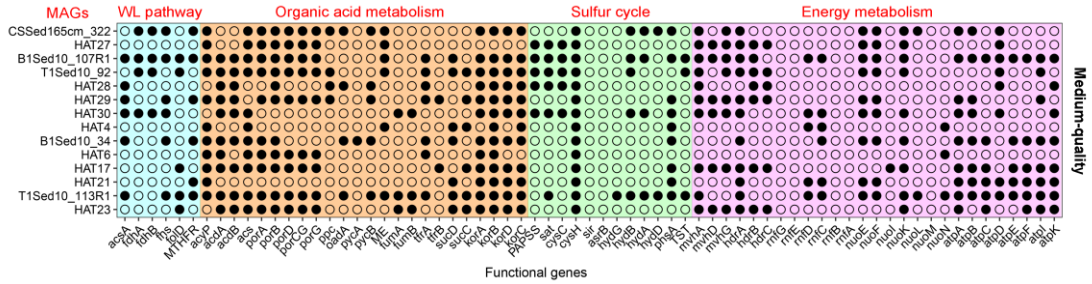

Fig. S8 Dot plot showing the detailed gene information involving in Wood-Ljungdahl pathway, organic acid metabolism, sulfur cycle and energy metabolism in medium-quality MAGs (including 9 MAGs obtained in this article and 5 downloaded from GTDB). Solid and hollow dots indicated the presence and absence in the MAGs, respectively. Enzymes and compounds abbreviations were listed in supplementary materials.

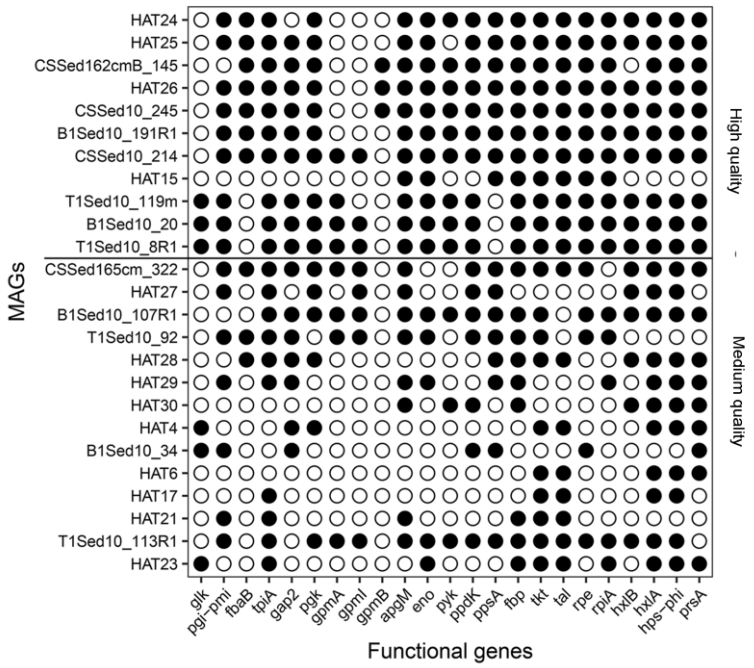

Fig. S9 Presence of genes involving in carbohydrate metabolism in *Ca.* Natranaeroarchaeales MAGs. The high- (completeness > 90%, contamination < 5%) and medium- quality (completeness > 70%, contamination < 10%) MAGs were included. Solid and hollow dots indicated the presence and absence in the MAGs, respectively. The definitions of genes were listed in Table S8.

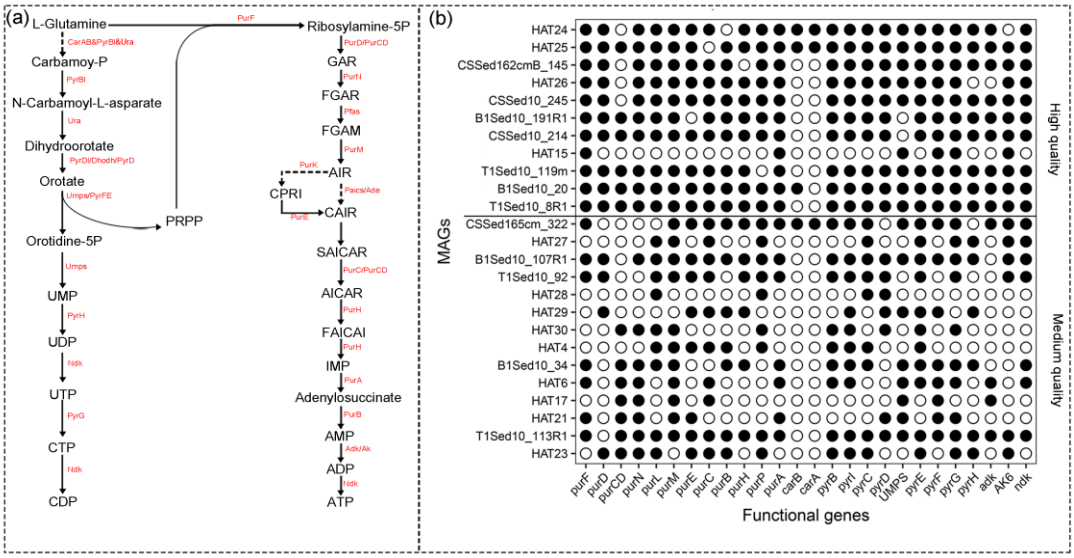

Fig. S10 Presence of genes involving in nucleotide biosynthesis in *Ca.* Natranaeroarchaeales MAGs. The high- (completeness > 90%, contamination < 5%) and medium- quality (completeness >70%, contamination < 10%) MAGs were included. Solid and hollow dots indicated the presence and absence in the MAGs, respectively. The definitions of genes were listed in Table S8, and enzymes and compounds abbreviations were shown in Supplementary Results.

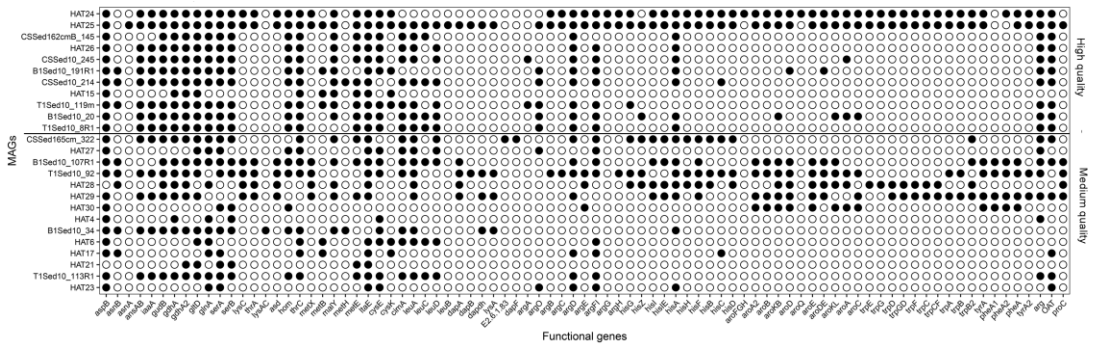

Fig. S11 Presence of genes involving in amino acid metabolism in *Ca.* Natranaeroarchaeales MAGs. The high- (completeness > 90%, contamination < 5%) and medium- quality (completeness > 70%, contamination < 10%) MAGs were included. Solid and hollow dots indicated the presence and absence in the MAGs, respectively. The definitions of genes were listed in Table S8.

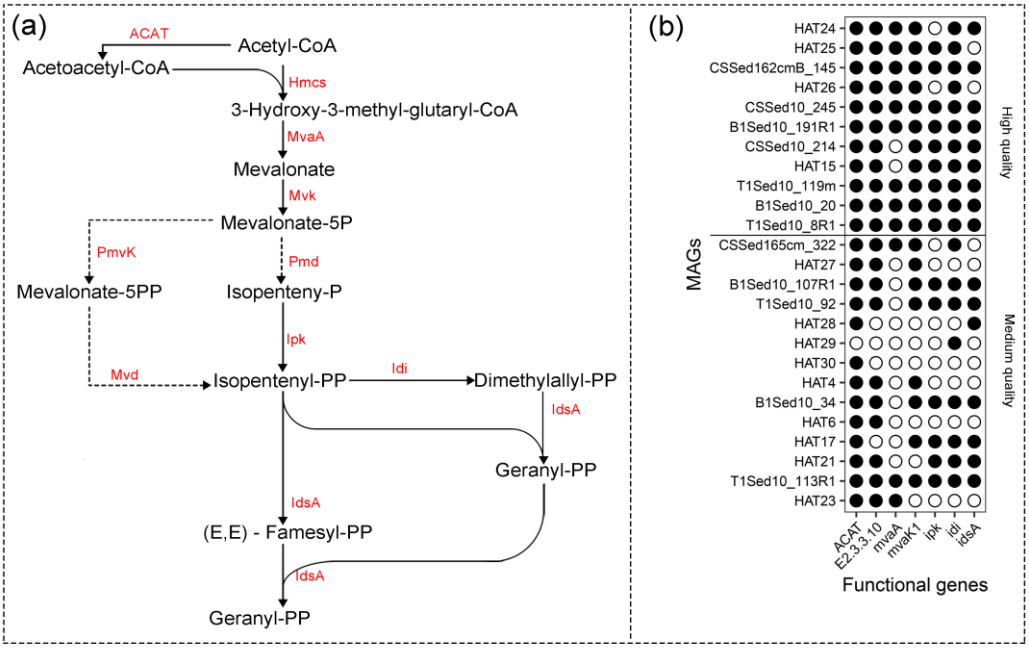

Fig. S12 Presence of genes involving in membrane biosynthesis in *Ca.* Natranaeroarchaeales MAGs. The high- (completeness > 90%, contamination < 5%) and medium- quality (completeness > 70%, contamination < 10%) MAGs were included. Solid and hollow dots indicated the presence and absence in the MAGs, respectively. The definitions of genes were listed in Table S8, and enzymes and compounds abbreviations were shown in Supplementary Results.

S

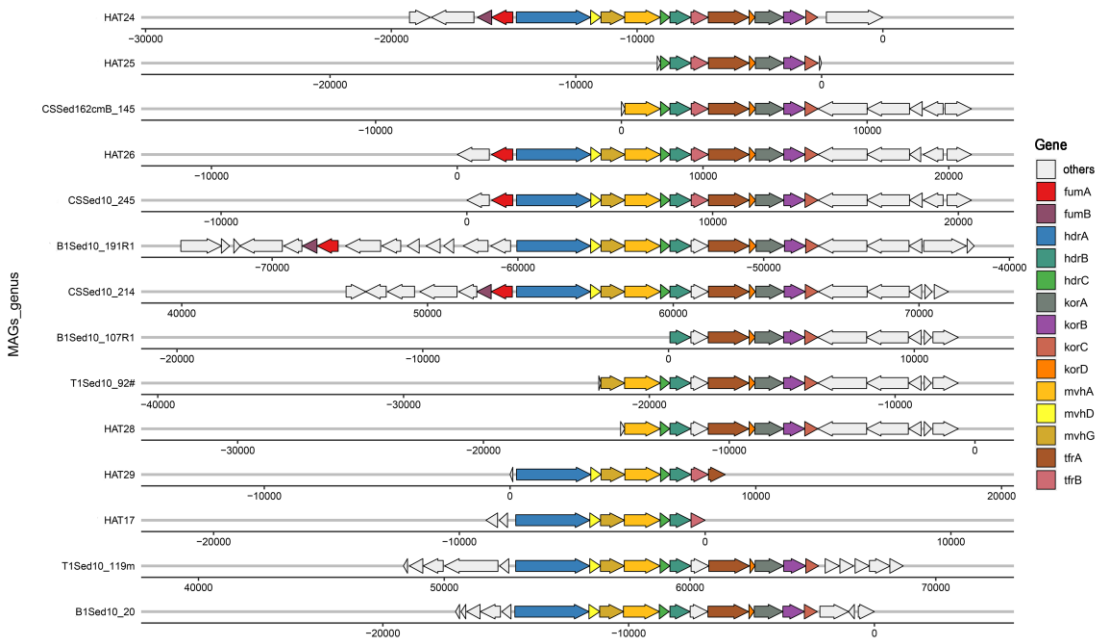

Fig. S13 Gene arrangement of *hdrABC*, *mvhADG*, *tfrAB* and korDABC in *Ca.* Natranaeroarchaeales. Longitudinal coordinate exhibited the MAG names. The genes were aligned by *hdrB*. The number in the horizontal ordinate indicated the position in the contig, and minus meant contig sequence was reverse complement. The length and direction of the arrow showed the gene length and direction. The fourteen interest genes were colored. The definitions of genes were listed in Table S8

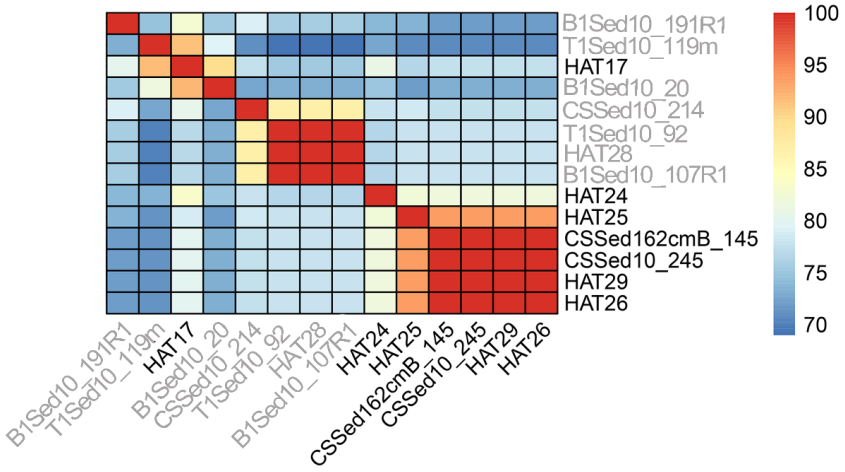

Fig. S14 Pairwise protein sequence alignment of TfrB and the homologs. The color indicated the identity value according to the scale. Labels indicated the MAGs names and grey marked the MAGs containing *tfrB* homolog genes.