Ecophysiological distinctions of haloarchaea from a hypersaline Antarctic lake determined using metaproteomics

Bernhard Tschitschko, Timothy J. Williams, Michelle A. Allen, Ling Zhong, Mark J. Raftery and Ricardo Cavicchioli

Supplemental material

MATERIALS AND METHODS

Sample collection from Deep Lake. Biomass was collected from Deep Lake (68°33'36.8S, 78°11'48.7E), Vestfold Hills, Antarctica between November 30 and December 5, 2008, by filtering water taken from 5, 13, 24 and 36 m depths (50 litres except 25 litres from 36 m), through a 20 µm prefilter sequentially onto 293 mm polyethersulfone membrane filters with 3.0, 0.8 and 0.1 µm pore sizes, as described previously (1). In addition, a surface sample (50 litres) was taken from close to the lake shore. All filters were placed in storage buffer, frozen in liquid nitrogen and cryogenically maintained at -80°C until being processed (1).

Metaproteomics. Protein extractions, mass spectrometry and data analysis were performed based on methods described previously (2), with minor modifications. Briefly, proteins were extracted without filtering protein suspensions through a 0.22 µm filter, protein concentrations were determined with the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, Rockford, IL, USA), and 25 µg protein of each sample was processed for mass spectrometry (MS) in a buffer containing a final concentration of 25 mM NH_4HCO_3, pH 8 – 8.5 (assessed using pH indicator strips). Samples were reduced with 2.5 mM dithiothreitol for 30 min at 37°C in the dark, alkylated with 5 mM iodoacetamide for 30 min at room temperature in the dark, trypsin digestion performed with 0.3 µg trypsin (Sequencing Grade

Modified Trypsin, Promega, Madison, WI, USA) overnight at 37°C, and peptides stored at -80°C until MS analysis. Peptide solutions were diluted 1:2 in 1% formic acid, 0.05% heptafluorobutyric acid and subjected to LC-MS/MS. LC-MS/MS was performed as described previously (3), with the only modification being the use of 3 μ, 200 Å instead of 5 μ, 200 Å C18 media in the nano column (Magic, Michrom Bioresources, Auburn CA, USA). At least two technical replicates were performed for each sample.

Peak lists from Orbitrap Velos mass spectra were generated using extract_msn and peptides identified through automated database searches using Mascot Daemon and the Mascot server (version 2.3; Matrix Science, Thermo, London, UK) with ThermoFinnigan LCQ/DECA RAW file as the import filter and the following settings: one accepted missed cleavage for the tryptic digest, peptide mass tolerance of +/- 4 p.p.m., fragment mass tolerance of +/- 0.4 Da and variable modifications of oxidation and carbamidomethylation. Ions were matched against peptides using a composite databases consisting of: 1) Deep Lake metagenome data, which comprised 5,837 contigs > 2 kb in length from the assembled contigs previously generated and annotated using SHAP representing 38,071 predicted protein sequences (1, 4) available at http://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=AntLakMetagenome (Antarctic Lakes Metagenome: whole_lake.gbk); and 2) all 14,181 predicted protein sequences from *Hht. litchfieldiae*, DL31, *Hrr. lacusprofundi* and DL1 sourced from the IMG portal (http://img.jgi.doe.gov/) (5). To facilitate calculations of false discovery rates (FDRs), the database contained randomized decoy proteins equal in number to those present in the reference database. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (6) with the dataset identifier PXD001436 and DOI 10.6019/PXD001436.

All technical replicates of a sample were combined and merged into a single sample file during Mascot analysis, resulting in 15 sample files (5 depths x 3 filters). Peptide and protein validation and further analysis was performed using Scaffold (version Scaffold_4.2.1, Proteome Software Inc., Portland, OR, USA) with strict settings of 95% minimum peptide identification probability and 99% minimum protein identification probability. The majority of proteins were identified with at least two peptide matches. The number of confidently identified proteins was maximized while minimizing false positives by including single peptide matches but maintaining low FDRs, as has been recommended (7, 8) and successfully applied in recent metaproteomic studies (9, 10, 11). The peptide and protein FDRs were 0.06% and 0.4%. For quantitative analyses the Normalized Total Spectrum Count (Scaffold_4.2.1, Proteome Software Inc., Portland, OR, USA) for each identified protein was combined across all 15 samples and used to quantify abundance.

All identified proteins were manually annotated using BLASTP searches against the genome encoded proteins of *Hht. litchfieldiae*, DL31, *Hrr. lacusprofundi*, and *Halobacterium* sp. DL1 on IMG, and against all entries in the ExPASy database (12), recording the best overall match including percentage identity, and organism. Annotated proteins were categorized into taxonomic and functional groups. Taxonomic categories were *Hht. litchfieldiae*; DL31; *Hrr. lacusprofundi*; DL1; Other *Halobacteriaceae* (best match to a member of the *Halobacteriaceae* other than the four Deep Lake isolates); Other *Archaea* (best match to a member of the *Archaea* other than *Halobacteriaceae*); *Bacteria*; Viruses; and *Dunaliella*. Proteins sharing the same set of detected peptides were grouped into protein families (2).

Functional categories were: Transport; Carbohydrate Metabolism; Glycerol Metabolism; Amino Acid Metabolism; Central Carbon Metabolism; Energy Conservation; Metabolism (Other); Cell Division; DNA Replication, Repair, and CRISPR; Oxidative Stress;

Transduction; Transcriptional Regulators; Transcription; Ribosomes; Protein Chaperones; Translation (Other); Proteolysis; Cell Surface; Hypothetical; Viruses. Cell Surface proteins were subdivided into subcategories based on proteins that comprise archaella (Archaella), adhesion pili (Adhesion), and the archaeal surface layer including hypothetical proteins that possess Sec, TAT or PGF-TERM sequences (Cell Surface Proteins – Other). The Hypothetical category included archaeal and bacterial proteins for which no function in the cell envelope or in cellular metabolism could be inferred. Hypothetical proteins were further subdivided into subcategories based on the presence of transmembrane helices (Hypothetical - Membrane); or nucleic-acid-binding domains (Hypothetical - Nucleic Acid Binding); or possessed domains which provided no indication as to function or possessed no identifiable domains at all (Hypothetical - Other). Other categories were subdivided into subcategories to provide increased resolution of cellular processes such as transport and metabolism. Transport: ABC Transporter - Amino Acids; ABC Transporter - Oligopeptides/Dipeptides; ABC Transporter - Carbohydrates; ABC Transporter - Phosphate/Phosphonate; ABC Transporter - Iron; ABC Transporter (Other); TRAP/TTT Transport; Cation Transport; Secretion; Other Transporter. Carbohydrate Metabolism: Glycosylation/Capsular Polysaccharide; Carbohydrate Metabolism (Other). Metabolism (Other): Nitrogen Metabolism; Sulfur Metabolism; Isoprenoid Metabolism; Vitamin/Cofactor Biosynthesis; Other (comprising those proteins inferred to be involved in metabolism, but no precise function or substrate specificity could be inferred based on identified domains). DNA Replication, Repair and CRISPR: DNA Replication and Repair; CRISPR.

Gene mapping. Genes in the tADL-II contigs were mapped onto the *Hht. litchfieldiae* tADL genome using the CONTIGuator web server (13) and by manual assignment, and circular plots were created using DNAPlotter (14) in Artemis (15).

Correlation analysis. Fractionation of the Deep Lake biomass according to depth and filter size allowed a comparison of the abundances (measured in spectrum counts) of single proteins and proteins from functional categories across samples thereby enabling ecological inferences to be made from observed co-expression. The analyses enabled statistically valid positive or negative correlations to be determined between spectrum counts for proteins across the 15 filter samples. Pair-wise comparisons of spectrum counts were made between individual abundant proteins, functional categories of proteins (the sum of the spectrum count for a functional group, such as tADL ABC transporter proteins) or taxonomic groups of proteins (for example all *Hht. litchfieldiae* tADL proteins vs all *Dunaliella* proteins). Pearson correlation coefficient and p-values between single proteins or protein functional categories across the 15 filter samples were calculated in R (16) with the Hmisc package (17) using normalized total spectrum counts. Only correlations with a p-value < 0.01 were regarded as statistically significant. All results of correlation analyses can be found in Table S2, and correlations mentioned in the manuscript are highlighted separately in Table S1.

Statistics using the PRIMER 6 software package. Statistical analyses were performed with PRIMER 6 **(**18**)** using the normalized total spectrum count for functional categories from each of the 15 samples. The data were standardized and square-root transformed prior to calculations using a Bray Curtis resemblance matrix. Non-metric multi-dimensional scaling (NMDS) plots were created using standard settings. Two-way crossed analysis of similarity (ANOSIM) without replicates were performed on the factors, filter size and sample depth, to test for statistically significant differences within these groups.

Epifluorescene microscopy. Deep Lake surface water samples were taken in the Antarctic summer of 2008/2009 and preserved in 2% (v/v) formaldehyde, or in 2013/2014 and preserved in 0.5% (v/v) glutaraldehyde, and all samples were stored at -80 °C until microscopy was performed. Deep Lake water (4 ml) was filtered through a 25 mm diameter,

0.02 µm pore size Whatman® Anodisc filter membrane (GE Healthcare Life Science, UK) with a 0.45 µm pore size backing filter (Type HA, Merck Millipore, MA, USA). Filters with captured biomass were stained with 10 µl SYBR® Gold nucleic acid stain (Invitrogen, Life Technologies, NY, USA) for 18 min in the dark and subsequently mounted on a glass slide with a drop of ProLong® Gold anti fade reagent (Invitrogen, Life Technologies, NY, USA). Microscopic analysis of slides was performed using an Olympus BX51WI epifluorescence microscope together with an Olympus DP71 camera and the cell Sense Standard imaging software (all Olympus, Hamburg, Germany). Slides were visualized under excitation with blue light (460 – 495 nm, emission 510 – 550 nm).

Growth studies. To assess growth characteristics based on inferences made from Deep Lake metaproteomic data, *Hht. litchfieldiae* was grown in batch cultures based on DBCM2 media (19, 20) using specific carbon and nitrogen sources. The substrates tested were DHA (10 mM) as a carbon source; starch (10 gL^{-1}) as a carbon source; acetamide (10 mM) as both a carbon source (with ammonia) or as a nitrogen source (with pyruvate); 2-aminoethylphosphonic acid (AEP; 5 mM) as a phosphorus source replacing phosphate in DBCM2 medium (tested both with and without peptone and yeast extract); *Hht. litchfieldiae* genomic DNA (200µg ml^{-1} final concentration) as a phosphorus source.

RESULTS AND DISCUSSION

Microscopy. To facilitate interpretation of the metaproteomic data, microscopy was performed on water samples to assess cell state (aggregated vs non-aggregated) and presence of particulate matter associated with 3.0, 0.8 and 0.1 µm filters. The samples preserved in glutaraldehyde provided clearer images than those preserved in formaldehyde (see Materials and Methods; data not shown). An increase in particulate matter was observed with increase in pore size, and cells were often associated with the particulate matter (Fig. S2).

149 ***Hht. litchfieldiae.*** Carbohydrate metabolism: The *Hht. litchfieldiae* genome encodes two
150 pathways for the conversion of glycerol into dihydroxyacetone phosphate (DHAP) (20). The
151 first catabolic pathway involves the ATP-dependent phosphorylation of glycerol to glycerol-
152 3-phosphate by glycerol kinase, followed by the oxidation of glycerol-3-phosphate to DHAP
153 by glycerol-3-phosphate dehydrogenase. The second catabolic pathway begins with the
154 oxidation of glycerol to DHA by glycerol dehydrogenase, followed by the phosphorylation of
155 DHA by DHA kinase to DHAP. DHAP is a versatile substrate that can enter into glycolysis
156 or gluconeogenesis; serve as a precursor for membrane phospholipids via glycerol-1-
157 phosphate; or be converted to methylglyoxal as a precursor for aromatic acid biosynthesis
158 (21) (see **Nitrogen and amino acid metabolism**). Evidence for the first glycerol catabolic
159 pathway included the detection of glycerol kinases and glycerol-3-phosphate dehydrogenase,
160 whereas for the second pathway DHA kinase subunits were detected, but no glycerol
161 dehydrogenase. The enzymes for the first pathway were also more abundant than for the
162 second pathway (Table S5). The lack of detection of glycerol dehydrogenase for *Hht.*
163 *litchfieldiae* might be due to repression of gene expression, which has been observed in *H.*
164 *salinarum* R1 as a possible mechanism to promote glycerol phosphorylation and decrease the
165 flow of glycerol to DHA (22). Thus, rather than being utilized for glycerol oxidation, it is
166 possible that DHA kinase is used solely to catabolize DHA directly obtained from the
167 environment. DHA is exuded as a byproduct of the breakdown of surplus glycerol in
168 *Dunaliella* (23, 24) and has been hypothesized to be an important growth substrate for
169 haloarchaea in hypersaline lakes (24).
170 The iron-containing glycerol dehydrogenase characterized for haloarchaea (25) has a
171 homolog in *Hht. litchfieldiae* (tADL_2148), but this was not detected in the metaproteome. It
172 is possible that a novel (for archaea) glycerol dehydrogenase may be present in the Deep
173 Lake metaproteome, including a glycerol dehydrogenase of the short-chain

dehydrogenase/reductase (SDH) family in bacteria (26), or a glycerol dehydrogenase of the aldo/keto reductase family in eukaryotes (27); *Hht. litchfieldiae* homologs of both were detected in the metaproteome, although their substrate specificity cannot be determined from the sequences.

Initially, no growth was detected for *Hht. litchfieldiae* using DHA as the defined carbon and energy source; but when *Hht. litchfieldiae* was grown on medium containing both DHA and pyruvate and was then transferred to medium containing just DHA, growth was observed (Fig. S5B). A similar response was observed previously for *Hht. litchfieldiae* where it grew with glycerol as the defined carbon and energy source only after being transferred from media containing both pyruvate and glycerol (20).

Phosphorus metabolism: AEP (ciliatine) was tested as a phosphorus source; this phosphonate and its N-alkylated derivatives are the most abundant and ubiquitous of naturally occurring phosphonates (28). Initially no growth of *Hht. litchfieldiae* was observed when AEP was tested as a phosphorus source in place of phosphate in the DBCM2 medium (lacking peptone and yeast extract). In addition to providing a phosphorus source, phosphate also serves to help buffer DBCM2 medium. The absence of phosphate buffer might therefore have an adverse effect on cell growth or viability. Although Tris.Cl was used to bring the pH of DBCM2 to 7.5 prior to the addition of AEP (5 mM), further Tris.Cl was added after the addition of AEP to reach a final concentration of ~5 mM and buffer at pH 7.5. This addition of Tris.Cl allowed *Hht. litchfieldiae* to grow on AEP as a source of phosphorus (Fig. S5A).

A putative DNA-binding membrane protein (halTADL_0044; winged helix-turn-helix DNA-binding domain) was detected for *Hht. litchfieldiae*. The encoding gene neighbors a gene (halTADL_0045) that is homologous to *H. volcanii* Hvo_1477, which is involved in the utilization of DNA as a phosphate source for growth (29). DNA concentrations can be particularly high in hypersaline environments (30) and the very low temperatures in Deep

Lake should further help to preserve extracellular DNA. Laboratory assessments indicated *Hht. litchfieldiae* was unable to utilize DNA as a phosphorus source for growth; this requires further evaluation in view of the metaproteome data.

Nitrogen and amino acid metabolism: An acetamidase homolog (amidohydrolase) was detected for *Hht. litchfieldiae* (halTADL_0419), but its function is unclear. *Hht. litchfieldiae* showed no growth in DBCM2 medium containing acetamide as a carbon source, which is consistent with the absence of genes that encode enzymes for acetate assimilation via either the methylaspartate cycle or glyoxylate cycle (20). However, *Hht. litchfieldiae* was also incapable of using acetamide as a nitrogen source in DBCM2 medium.

The detection of a putative copper-containing nitrite reductase (nitric oxide forming; halTADL_2997) and halocyanin acceptor protein (halTADL_2996) provides possible evidence of active denitrification by *Hht. litchfieldiae*. The use of nitrite as an electron acceptor might indicate oxygen depleted conditions, at least in the micro-environment of the *Hht. litchfieldiae* cells at the time of sampling.

The detection of S-adenosylmethionine (SAM) synthetase (halTADL_3028) indicates that methionine is converted to SAM, an important methyl donor for processes including DNA methylation and cofactor biosynthesis (heme, cobalamin (31)). The product of SAM demethylation (S-adenosyl-homocysteine) can be hydrolyzed to homocysteine; the enzyme responsible (S-adenosylhomocysteinase, halTADL_1723) was also detected. Further, in archaea homocysteine is a precursor to cysteine via the path leading through cystathionine (25, 32), which might suggest that cysteine is synthesized from methionine by *Hht. litchfieldiae* in Deep Lake. We also detected three proteins corresponding to enzymes involved in the generation of chorismate from fructose-1,6-bisphosphate(halTADL_0575, halTADL_0574, halTADL_2582), as well as enzymes specific to the synthesis of tryptophan

(halTADL_0576, halTADL_3066, halTADL_0889) and phenylalanine (halTADL_2073) (33), indicating aromatic acid synthesis in Deep Lake.

Motility and taxis: The *Hht. litchfieldiae* MCPs included HtrII for sensory rhodopsin II (SRII; phoborhodopsin), which is indicative of negative phototaxis as the repellent receptor SRII is activated by blue light and is produced when respiratory activity is high and cells seek darker conditions to minimize photo-oxidative damage (34) (Also see main text section **Haloarchaeal responses to Antarctic solar irradiation,** and Table S7). One MCP belongs to the HemAT family of heme-based transducers involved in aerotaxis (35), and another has an N-terminal globin domain and may function in aerotaxis or oxidative stress. The MCP halTADL_1218 has similarity to MCPs involved in chemotaxis of organic nutrients (36), and consistent with this, the gene is located within a cluster of genes which function to uptake and catabolize lactate.

One-carbon (C1) metabolism: The major C1 carrier in haloarchaea is tetrahydrofolate (H_4F), which is essential for the biosynthesis of methionine, glycine, purines, and thymidine (37). Several proteins involved in the biosynthesis of this cofactor were detected in the metaproteome, as well as H_4F-dependent proteins (e.g., purine biosynthesis in *Hht. litchfieldiae*). However, there was evidence of expression of a methanopterin-based C1 transfer system in *Hht. litchfieldiae* in Deep Lake, with tetrahydromethanopterin (H_4MPT) used as the carrier. The H_4MPT-associated C1 transfer enzyme methenyl-H_4MPT cyclohydrolase (Mch; halTADL_3392) was detected in the metaproteome, as well as a NADPH-dependent F_{420} reductase (halTADL_2320). Both H_4MPT and cofactor F_{420} were likely inherited from methanogenic archaea, given the posited origin of haloarchaea (38, 39). The function of the H_4MPT system in *Hht. litchfieldiae* is unclear; there are no clear homologs of genes for either methylotrophy or formaldehyde activation in the genome, and it is not known this H_4MPT-dependent pathway proceeds in the oxidative or reductive

directions. For example, the sulfate-reducing archaeon *Archaeoglobus fulgidus* cleaves acetyl-CoA into a methyl group and CO_2, and subsequent oxidation of the H_4MPT-bound methyl group serves as a source of energy (40); but this pathway is unlikely for *Hht. litchfieldiae* in the absence of genes encoding a complete carbon monoxide dehydrogenase/acetyl-CoA complex required for acetyl-CoA cleavage. One possibility is that methyl groups are generated internally via catabolism (such as methylphosphonate; see **Phosphorus metabolism**), and either oxidized for energy or assimilated by *Hht. litchfieldiae*. Genes for certain H_4MPT-associated C1 transfer proteins (including methylene-H_4MPT reductase) are found adjacent to the gene cluster for phosphonate degradation in the *Hht. litchfieldiae* genome (20). If *Hht. litchfieldiae* utilizes phosphonates as a carbon source, then C1 compounds might be transferred to H_4MPT and oxidized to the formyl level; this latter step could be catalyzed by Mch for liberated methyl groups. It is possible that the Mch product (formyl-H_4MPT) serves a biosynthetic function; but (unlike formyl-H_4F) formyl-H_4MPT has been regarded as unsuited to formyl donations for biosynthesis (41).

Heavy metal efflux: The concentrations of lead (3.7-5.2 µg/L) and copper (9.1 to 21 µg/L; copper concentrations showed a marked decrease with depth) in Deep Lake were found to be much higher than seawater, although not at levels high enough to inhibit microbial growth (42). This may account for a putative Cu^{2+}-exporting ATPase (halTADL_1767) detected for *Hht. litchfieldiae*.

***Hht. litchfieldiae* variants**. GC/read-depth profiling of the *de novo* assembly of metagenome data revealed a cluster of 52 large contigs (>15 kb) totalling 1.89 Mb which had highest identity (~85%) to *Hht. litchfieldiae* tADL, and the genes on the contigs tended to be syntenic with tADL (1). The contig cluster was referred to as the 'tADL-related 5th genome' (1), and has now been designated *Hht. litchfieldiae* strain 'tADL-II'. A total of 107 proteins

were identified from 38 of the contigs (Table S9). The proteins on the contigs that matched tADL had ~70 – 99% sequence identity (lower for some cell surface proteins).

Thirty eight additional variants on 22 contigs were further assigned to tADL-II (Table S9) even though the corresponding contigs were not part of the original GC/read-depth binning. These variants are derived from contigs each containing multiple ORFs with ~70 – 99% amino acid identity to tADL sequences (lower for some cell surface proteins) and exhibiting gene synteny with tADL; characteristics shared with the contigs previously assigned to tADL-II. These included 11 ribosomal proteins that had 88 – 97% sequence identity to tADL. A total of 18 of the 22 contigs were shorter than 15 kb and would have been excluded from the previous analyses (1). Scaffolding of the 52 original contigs with the 22 new contigs revealed that 10 of the new contigs overlapped (at their ends) with the original contigs and were assembled into larger contigs. These 10 contigs encoded for 24 of the newly assigned tADL-II proteins. Mean read-depth/GC-content was also comparable with 38.8/63% for the original set of 52 contigs and 35/61% for the additional 22 contigs.

High variation was observed between tADL and tADL-II sequences for cell surface proteins (63% average amino acid identity) in contrast to substantially less variation for typically conserved, transcription and cell division proteins (≥94%) (Fig. 4). The divergence between tADL and tADL-II proteins is likely to result in phenotypic distinctions, as even single amino acid changes can confer functional differences (e.g. in the active-site, substrate binding site, site of interaction for effector molecules, protein-protein interactions). Other proteomic distinctions between tADL and tADL-II included proteins encoded by genes present on tADL-II contigs that were absent in the tADL genome. The detection of a nitrate/sulfonate/bicarbonate ABC transporter solute-binding lipoprotein (unique to tADL-II) may confer the ability to target distinct nutrient sources (Table S9).

In contrast to the level of variation for contigs assigned to tADL-II, proteins were identified matching to contigs which had overall high identity to tADL. These contigs had ORFs with 100% identity to tADL plus some ORFs (typically one) with 97-99% identity, and were therefore assigned as variants of tADL. These comprised a total of 8 variants for which 6 had 99% identity (5 single SNPs plus 1 with a 3 nt deletion) and 2 had 97% identity, and all had neighbouring genes with 100% sequence identity and conserved gene synteny with tADL.

One of the variants was an ABC transporter phosphate-binding lipoprotein which arose from a previously identified SNP (1). The SNPs characterized in the previous study represented ≥90% of the population (1), indicating it was the dominant form in the population. In addition, two other variants of the same protein (88% and 93% sequence identity) were identified in the metaproteome, both of which were encoded on contigs that could potentially be assigned to tADL-II. The amino acid changes for the transporter may confer functional differences in *Hht. litchfieldiae* phosphate acquisition. Multiple cell surface protein variants were previously identified which mapped to small regions of the tADL genome that had low fragment recruitment of metagenome data, with the high level of variation thought to provide a mechanism for hosts to evade viral infection (2).

In another case, one α-amylase was detected with 94% sequence identity to tADL and the only other ORF on the contig matched tADL with 96% sequence identity, and the two ORFs were syntenic with tADL. For the same α-amylase, a protein with 100% match to tADL was also identified (therefore assigned to tADL), and one that matched to tADL-II (84% identity to tADL). The data suggest that the 94% match could therefore derive from a variant of *Hht. litchfieldiae*. However, intergenera gene exchange has also been documented for the Deep Lake community, and it is possible that the contig derives from one of the other haloarchaeal species in the lake. From a functional standpoint, similar to the ABC transporter

phosphate-binding lipoprotein (above), the extent of variation of the α-amylase proteins may increase the capacity of the lake population to utilize starch-related substrates.

For a total of 15 proteins, the best match was to *Hht. litchfieldiae*, but other ORFs on the contigs matched to other haloarchaea or viruses. These included three distinct proteins, all matching to the same *Hht. litchfieldiae* glycerol kinase with 98% sequence identity. These findings are consistent with the dissemination of genes within the lake population and selection for a collective genetic capacity to effectively exploit available nutrients throughout the lake.

DL31 and *Hrr. lacusprofundi* variants. Only four respectively three variants were detected for DL31 and *Hrr. lacusprofundi*, with five of them associated with the haloarchaeal cell surface (Table S10). One of the *Hrr. lacusprofundi* variants represents an archellin protein with a particularly high level of variation (38%). An additional archaellin protein with less variation to *Hrr. lacusprofundi* (77%) mapped to a contig where the neighboring genes matched best to *Halorubrum* spp., but not specifically *Hrr. lacusprofundi* (Table S11). These findings provide further support for the importance of archaellin variation within the *Hrr. lacusprofundi* population (2). In addition, two variants matched to DL31 but neighboring genes matched to other haloarchaea (Table S11). Four variants, including a carbohydrate ABC transporter protein matched to *Hrr. lacusprofundi* but with neighboring contig genes matching to other haloarchaea (Table S11). The contigs matching other haloarchaea may represent islands of genomic DNA present within strains of *Hrr. lacusprofundi*, or derive from other low abundance haloarchaeal species in the lake.

DL31 metabolism and cell function. DL31 oligopeptide and iron ABC transporters were positively correlated with the highly abundant DL31 protein Halar_1791 of unknown function (Table S1). DL31 protein Halar_1791 was the 5th most abundant protein for DL31 (29th overall in the metaproteome). It belongs to the MmpL (mycobacterial membrane protein

large) transporter family of the extended RND (Resistance-Nodulation-Cell Division) permease superfamily (43, 44). Thus, we infer that Halar_1791 is a transporter. Homologs of Halar_1791 are found across *Halobacteriales* genomes, although their function has not been experimentally determined. In mycobacteria, MmpL transporters are involved in lipid export and antibiotic efflux (43, 45, 46). The distribution of Halar_1791 protein was positively correlated with the two most abundant proteins for DL31, oligopeptide- and iron-binding lipoproteins of ABC transporters, and we therefore speculate that the three transporters may be functionally associated. Halar_1791 also showed a positive correlation with most *Hht. litchfieldiae* archaellin proteins (Tables S1 and S2). It also showed a positive correlation with the most abundant *Hrr. lacusprofundi* TRAP transporter (Hlac_2586) and oligopeptide ABC transporter (Hlac_0069), and bacterial porin and TonB transporter receptor. Overall, these correlations underscore the potential importance of Halar_1791 to the acquisition of nutrients by DL31. Another MmpL/RND permease transporter was detected in the Deep Lake metaproteome for *Hht. litchfieldiae* (halTADL_0082), but at very low abundance.

General features of *Hht. litchfieldiae*, DL31 and *Hrr. lacusprofundi* in Deep Lake.

Cell surface and glycosylation: The surface layer (S-layer) glycoprotein forms a paracrystalline lattice that functions as the haloarchaeal cell envelope (2, 47). Detected putative cell surface glycoproteins of the three dominant haloarchaea (halTADL_1043, Halar_0829, Hlac_2976, Hlac_0412) are likely to be the major S-layer proteins, which accounts for their high abundance in the samples. These form a porous, two-dimensional lattice that encases the cell, with the S-layer proteins anchored in the cell membrane, and the S-layer envelope separated from the cell membrane by a ‘quasi-periplasmic space’ (47). In addition, S-layer proteins and other surface-exposed proteins (archaella, pilins) are post-translationally modified by glycosylation (48, 49, 50, 51). N-glycosylation of S-layer proteins proceeds via a process involving multiple archaeal glycosylation (Agl) proteins, of which the

most conserved is oligosaccharyltransferase AglB, a membrane-bound protein that transfers the glycan moiety to the protein (50). Aside from AglB, little if any sequence identity is shared among other glycosylation enzymes across haloarchaeal species (50). AglB was detected for *Hht. litchfeldiae* (halTADL_2411). Other proteins in the metaproteome were implicated in glycosylation of surface proteins, based on homologs involved in N-glycosylation of proteins in other haloarchaea (49, 50, 52); these include a sugar nucleotidyltransferase (halTADL_3353), a glycosyltransferases (halTADL_2565), NUDIX hydrolases (halTADL_3253, halTADL_2550), and a nucleoside-diphosphate-sugar epimerase/hydratase (halTADL_3057).

Isoprenoid metabolism: The cell membrane lipids of haloarchaea consist of glycerol diether lipids with prenyl side chains. The glycerophosphate backbone of these membrane lipids is formed from glycerol 1-phosphate (see **Carbohydrate metabolism**), and the prenyl side chains and other isoprenoids are derived via the mevalonate pathway (25, 53). The isoprene-based lipids used for lipid modification of S-layer glycoproteins in haloarchaea are also synthesized via the mevalonate pathway (54). Isoprenoid biosynthesis proteins were detected in the metaproteome, including an enzyme (phytoene synthase; halTADL_0465) involved in retinal production for rhodopsins in *Hht. litchfieldiae*.

Sulfur metabolism: The mechanism for reductive sulfate assimilation in haloarchaea has yet to be resolved (25, 55). Several hypothetical rhodanese domain proteins have been proposed as thiosulfate sulfurtransferase or sulfite reductase in haloarchaea (55); homologs of these proteins were detected for the three dominant Deep Lake haloarchaea (halTADL_2750, Halar_2935, Hlac_1687). *Hht. litchfieldiae* has been reported to produce H_2S from thiosulfate (56). A detected sulfotransferase (halTADL_1176) showed closest sequence identity to phosphoadenylylsulfate (PAPS) reductase; PAPS is the universal sulfate donor for the

sulfation of lipids and sugars (57), so this enzyme may function in sulfate transfers required for the biosynthesis of sulfated lipids and/or sugars.

Osmotic adaptation: A PspA homolog was detected for *Hht. litchfieldiae* (halTADL_2278). In bacteria, the transcriptional activator PspA plays a role in sensing a variety of membrane stressors, including phage infection, heat shock, and osmotic shock (58). Proteomic analysis of *H. volcanii* showed that the PspA homolog was more abundant in high-salt conditions, suggesting that it may play an important role in hypersaline adaptation, although the mechanism is not known (59). In light of virus infection of *Hht. litchfieldiae* in Deep Lake (2), it is also possible that the PspA homolog in *Hht. litchfieldiae* is also responsive to membrane perturbation resulting from infection.

Homologs of osmotic inducible protein C (OsmC) were detected for DL31 (Halar_1442) and *Hrr. lacusprofundi* (Hlac_1348). OsmC was originally identified in *E. coli* in response to osmotic shock (60), and accumulates in cells exposed to high external osmolality (61). Because OsmC displays peroxiredoxin-like activity against organic hydroperoxides, it has been inferred to have cross-protectivity against elevated osmolality and oxidative stress (61). However, in a proteomic study of the archaeon *Thermococcus kodakaraensis*, OsmC was found to have increased abundance in response to osmotic stress, but not oxidative stress (62).

Adhesion: Distinct from archaella, type IV pilus proteins (PilA) were detected for *Hht. litchfieldiae* (halTADL_1387, halTADL_0751, halTADL_1387), DL31 (Halar_2365, Halar_3709), and *Hrr. lacusprofundi* (Hlac_1363, Hlac_3311) (Table S3). These pili are essential for some haloarchaea to adhere to surfaces (48, 63, 64), and a large DL31 pilus protein (653 aa) contained a PKD domain which may promote intercellular interactions in archaea (65).

Cell division and growth: Haloarchaea typically have a FtsZ-based system for cytokinesis (66). FtsZ proteins were detected in the Deep lake metaproteome, matching *Hht. litchfieldiae* (halTADL_0937, halTADL_3056) and DL31 (Halar_2224). A possible MreB/FtsA cell division protein was also detected for *Hht. litchfieldiae* (halTADL_0130). Another protein implicated in cell division is an ATPase containing dual CDC48 (Cell division control protein 48) domains, which belongs to a class of VCP (Valosin Containing Protein)-like archaeal proteins that are implicated in the regulation of the cell cycle (67); these were detected for *Hht. litchfieldiae* (halTADL_2740) and DL31 (Halar_1865, Halar_2098).

Energy metabolism: ATP synthase and/or respiratory chain proteins were detected for *Hht. litchfieldiae*, DL31, and *Hrr. lacusprofundi* consistent with the generation of metabolic energy. The detection of a cytoplasmic inorganic pyrophosphatase for *Hht. litchfieldiae* (halTADL_1644) is undoubtedly important to energy metabolism, given that pyrophosphate is generated as a byproduct of numerous metabolic processes (including phosphonate degradation; see **Phosphorus metabolism**), and pyrophosphate hydrolysis is a highly exergonic reaction that can be used to facilitate less energetically favourable processes (68).

Other microorganisms in Deep Lake. *Dunaliella*: It is likely that the lake's primary producer *Dunaliella* is underrepresented in the Deep Lake metaproteome data as metagenome analyses identified few matches to available *Dunaliella* sequence data (1). In the metaproteome, a total of six chloroplast proteins were detected: a translation initiation factor, ribulose bisphosphate carboxylase/oxygenase (large chain), and chloroplast ATP synthase subunits. These provide evidence for photosynthesis and carbon fixation by *Dunaliella*.

Halobacterium sp. DL1: Proteins were detected for *Halobacterium* sp. DL1, a low abundance (~0.3% of the community (20)) haloarchaeaon in Deep Lake, including S-layer glycoprotein, ATP synthase subunit, an archaellin, and a BCAA ABC transporter lipoprotein.

444 The expression of this last protein is consistent with genomic evidence, by which DL1 was
445 inferred to have a metabolic preference for amino acids, especially BCAA (20).
446 Bacteria: Thirty six proteins showed the best matches to bacterial proteins, although
447 these were typically low matches, with half showing less than 50% sequence identity to
448 known bacterial proteins. Most of the proteins showed the best matches to
449 *Gammaproteobacteria*, including *Alteromonadales* (especially *Marinobacter* spp.) and
450 *Oceanospirillales*; based on SSU 16S rRNA pyrotag data, these groups have been detected in
451 Deep Lake (20). Putative bacterial proteins include cell surface proteins involved in
452 intercellular interactions; an outer membrane conjugative transfer protein (*Marinobacter* sp.
453 TraF (69)); and a TonB-dependent receptor and a periplasmic component of a TRAP
454 transporter, which indicate targeting of complex substrates and carboxylates, respectively.
455
456 **REFERENCES**

457 1. **DeMaere MZ, Williams TJ, Allen MA, Brown MV, Gibson JA, Rich J, Lauro FM,**
458 **Dyall-Smith M, Davenport KW, Woyke T, Kyrpides NC, Tringe SG, Cavicchioli R.**
459 2013. High level of intergenera gene exchange shapes the evolution of haloarchaea in an
460 isolated Antarctic lake. Proc Natl Acad Sci USA **110:** 16939–16944.
461 2. **Tschitschko B, Williams TJ, Allen MA, Páez-Espino D, Kyrpides N, Zhong L,**
462 **Raftery MJ, Cavicchioli R.** 2015. Antarctic archaea-virus interactions: metaproteome-
463 led analysis of invasion, evasion and adaptation. ISME J **9:** 2094–2107.
464 3. **Williams TJ, Long E, Evans F, DeMaere MZ, Lauro FM, Raftery MJ, Ducklow H,**
465 **Grzymski JJ, Murray AE, Cavicchioli R.** 2012. A metaproteomic assessment of winter
466 and summer bacterioplankton from Antarctic Peninsula coastal surface waters. ISME J **6:**
467 1883-1900.

4. **DeMaere MZ, Lauro FM, Thomas T, Yau S, Cavicchioli R.** 2011. Simple high-throughput annotation pipeline (SHAP). Bioinformatics **27:** 2431-2432.
5. **Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Anderson I, Lykidis A, Mavromatis K, Ivanova NN, Kyrpides NC.** 2010. The integrated microbial genomes system: an expanding comparative analysis resource. Nucleic Acids Res **38:** D382-390.
6. **Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, Dianes JA, Sun Z, Farrah T, Bandeira N, Binz PA, Xenarios I, Eisenacher M, Mayer G, Gatto L, Campos A, Chalkley RJ, Kraus HJ, Albar JP, Martinez-Bartolomé S, Apweiler R, Omenn GS, Martens L, Jones AR, Hermjakob H.** 2014. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotech **32:** 223-226.
7. **Gupta N, Pevzner PA.** 2009. False discovery rates of protein identifications: a strike against the two-peptide rule. J Proteome Res **8:** 4173-4181.
8. **Claassen M.** 2012. Inference and validation of protein identifications. Mol Cell Prot **11:** 1097-1104.
9. **Morris RM, Nunn BL, Frazar C, Goodlett DR, Ting YS, Rocap G.** 2010. Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. ISME J **4:** 673-685.
10. **Schneider T, Keiblinger KM, Schmid E, Sterflinger-Gleixner K, Ellersdorfer G, Roschitzki B, Richter A, Eberl L, Zechmeister-Boltenstern S, Riedel K.** 2012. Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. ISME J **6:** 1749-1762.

11. **Herbst FA, Bahr A, Duarte M, Pieper DH, Richnow HH, von Bergen M, Seifert J, Bombach P.** 2013. Elucidation of in situ polycyclic aromatic hydrocarbon degradation by functional metaproteomics (protein-SIP). Proteomics **13:** 2910-2920.

12. **Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Morreti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H.** 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res **40:** W597-603.

13. **Galardini M, Biondi EG, Bazzicalupo M, Mengoni A.** 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. Source Code Biol Med **6:** 11.

14. **Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J.** 2009. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics **25:** 119-120.

15. **Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA.** 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics **28:** 464-469.

16. **R Core Team.** 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

17. **Frank E Harrell Jr, with contributions from Charles Dupont and many others.** 2014. Hmisc: Harrell Miscellaneous. R package version 3.14-6. http://CRAN.R-project.org/package=Hmisc

18. **Clarke KR, Gorley RN.** 2006.PRIMER v6: User Manual/Tutorial. Primer-E, Plymouth, 192pp.

19. **Burns DG, Dyall Smith M.** 2006. Cultivation of haloarchaea. Methods Microbiol **35:** 535–552.

20. **Williams TJ, Allen MA, DeMaere MZ, Kyrpides NC, Tringe SG, Woyke T, Cavicchioli R.** 2014. Microbial ecology of an Antarctic hypersaline lake: genomic assessment of ecophysiology among dominant haloarchaea. ISME J **8:** 1645–1658.

21. **Oren A, Gurevich P.** 1995. Diversity of lactate metabolism in halophilic archaea. Can J Microbiol **41:** 302-307.

22. **Schwaiger R, Schwarz C, Furtwängler K, Tarasov V, Wende A, Oesterhelt D.** 2010. Transcriptional control by two leucine-responsive regulatory proteins in *Halobacterium salinarum* R1. BMC Mol Biol **11:** 40.

23. **Elevi Bardavid R, Oren A.** 2008. Dihydroxyacetone metabolism in *Salinibacter ruber* and in *Haloquadratum walsbyi*. Extremophiles **12:** 125–131.

24. **Ouellette M, Makkay AM, Papke RT.** 2013. Dihydroxyacetone metabolism in *Haloferax volcanii*. Front Microbiol **4:** 376.

25. **Falb M, Müller K, Königsmaier L, Oberwinkler T, Horn P, von Gronau S, Gonzales O, Pfeiffer F, Bornberg-Bauer E, Oesterhelt D.** 2008. Metabolism of halophilic archaea. Extremophiles **12:** 177-196.

26. **Monniot C, Zébré AC, Aké FM, Deutscher J, Milohanic E.** 2012. Novel listerial glycerol dehydrogenase- and phosphoenolpyruvate-dependent dihydroxyacetone kinase system connected to the pentose phosphate pathway. J Bacteriol **194:** 4972-4982.

27. **Jung JY, Kim TY, Ng CY, Oh MK.** 2012. Characterization of GCY1 in *Saccharomyces cerevisiae* by metabolic profiling. J Appl Microbiol **113:** 1468-1478.

28. **Hildebrand RL.** 1983. The effects of synthetic phosphonates on living systems. In: Hildebrand RL (ed). The Role of Phosphonates in Living Systems. CRC Press, Inc., Boca Raton, pp 139–169

29. **Chimileski S, Dolas K, Naor A, Gophna U, Papke RT.** 2014. Extracellular DNA metabolism in *Haloferax volcanii*. Front Microbiol **5:** 57.

30. **Danovaro R, Corinaldesi C, Dell'Anno A, Fabiano M, Corselli C.** 2005. Viruses, prokaryotes and DNA in the sediments of a deep-hypersaline anoxic basin (DHAB) of the Mediterranean Sea. Environ Microbiol **7:**586-592.

31. **Buchenau B, Kahnt J, Heinemann IU, Jahn D, Thauer RK.** 2006. Heme biosynthesis in *Methanosarcina barkeri* via a pathway involving two methylation reactions. J Bacteriol **188:** 8666-8668.

32. **White RH.** 2003. The biosynthesis of cysteine and homocysteine in *Methanococcus jannaschii*. Biochim Biophys Acta **1624:** 46–53.

33. **Gulko MK, Dyall-Smith M, Gonzalez O, Oesterhelt D.** 2014. How do haloarchaea synthesize aromatic amino acids? PLoS One **9:**e107475.

34. **Spudich JL.** 2006. The multitalented microbial sensory rhodopsins. Trends Microbiol **14:** 480-487.

35. **Hou S, Larsen RW, Boudko D, Riley CW, Karatan E, Zimmer M, Ordal GW, Alam M.** 2000. Myoglobin-like aerotaxis transducers in Archaea and Bacteria. Nature **403:** 540-544.

36. **Storch KF, Rudolph J, Oesterhelt D.** 1999. Car: a cytoplasmic sensor responsible for arginine chemotaxis in the archaeon *Halobacterium salinarum*. EMBO J **18:** 1146-1158.

37. **Levin I, Giladi M, Altman-Price N, Ortenberg R, Mevarech M.** 2004. An alternative pathway for reduced folate biosynthesis in bacteria and halophilic archaea. Mol Microbiol **54:**1307-1318.

38. **Nelson-Sathi S, Dagan T, Landan G, Janssen A, Steel M, McInerney JO, Deppenmeier U, Martin WF.** 2012. Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. Proc Natl Acad Sci USA **109:** 20537-20542.

565 39. **Nelson-Sathi S, Sousa FL, Roettger M, Lozada-Chávez N, Thiergart T, Janssen A,**
566 **Bryant D, Landan G, Schönheit P, Siebers B, McInerney JO, Martin WF.** 2015.
567 Origins of major archaeal clades correspond to gene acquisitions from bacteria. Nature
568 **517:** 77-80.
569 40. **Möller-Zinkhan D, Börner G, Thauer RK.** 1989. Function of methanofuran,
570 tetrahydromethanopterin, and coenzyme F_{420} in *Archaeoglobus fulgidus*. Arch Microbiol
571 **152:** 362-368.
572 41. **Maden BEH.** 2000. Tetrahydrofolate and tetrahydromethanopterin compared:
573 functionally distinct carriers in C_1 metabolism. Biochem J **350:** 609-629.
574 42. **Barker RJ.** 1981. Physical and chemical parameters of Deep Lake, Vestfold Hills,
575 Antarctica. Publ. No. 130. Australian National Antarctic Research Expeditions Series
576 B(V) Limnology, 73 pp.
577 43. **Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG, Cole ST.** 1999. Analysis of
578 the proteome of *Mycobacterium tuberculosis* in silico. Tuber Lung Dis **79:** 329–342
579 44. **Sandhu P, Akhter Y.** 2015. The internal gene duplication and interrupted coding
580 sequences in the MmpL genes of *Mycobacterium tuberculosis*: Towards understanding
581 the multidrug transport in an evolutionary perspective. Int J Med Microbiol **305:** 413–
582 423.
583 45. **Converse SE, Mougous JD, Leavell MD, Leary JA, Bertozzi CR, Cox JS.** 2003.
584 MmpL8 is required for sulfolipid-1 biosynthesis and *Mycobacterium tuberculosis*
585 virulence. Proc Natl Acad Sci USA **100:** 6121–6126.
586 46. **Pacheco SA, Hsu FF, Powers KM, Purdy GE.** 2013. MmpL11 protein transports
587 mycolic acid-containing lipids to the mycobacterial cell wall and contributes to biofilm
588 formation in *Mycobacterium smegmatis*. J Biol Chem **288:** 24213–24222.
589 47. **Albers S-J, Meyer BH.** 2011. The archaeal cell envelope. Nat Rev Microbiol **9:** 414-426.

590 48. **Tripepi M, Imam S, Pohlschroder M.** 2010. *Haloferax volcanii* flagella are required for
591 motility but are not involved in PibD-dependent surface adhesion. J Bacteriol **192:** 3093-
592 3102.
593 49. **Kaminski L, Guan Z, Yurist-Doutsch S, Eichler J.** 2013a. Two distinct N-
594 glycosylation pathways process the *Haloferax volcanii* S-layer glycoprotein upon changes
595 in environmental salinity. MBio **4:** e00716–13.
596 50. **Kaminski L, Lurie-Weinberger MN, Allers T, Gophna U, Eichler J.** 2013b.
597 Phylogenetic- and genome-derived insight into the evolutionary history of N-
598 glycosylation in Archaea. Mol Phylogenet Evol **68:** 327–339.
599 51. **Jarrell KF, Ding Y, Meyer BH, Albers SV, Kaminski L, Eichler J.** 2014. N-linked
600 glycosylation in Archaea: a structural, functional, and genetic analysis. Microbiol Mol
601 Biol Rev **78:** 304-341.
602 52. **Kaminski L, Eichler J.** 2014. *Haloferax volcanii* N-glycosylation: delineating the
603 pathway of dTDP-rhamnose biosynthesis. PLoS One **9** :e97441.
604 53. **VanNice JC, Skaff DA, Wyckoff GJ, Miziorko HM.** 2013. Expression in *Haloferax*
605 *volcanii* of 3-hydroxy-3-methylglutaryl coenzyme A synthase facilitates isolation and
606 characterization of the active form of a key enzyme required for polyisoprenoid cell
607 membrane biosynthesis in halophilic archaea. J Bacteriol **195:** 3854-3862.
608 54. **Konrad Z, Eichler J.** 2002. Lipid modification of proteins in Archaea: attachment of a
609 mevalonic acid-based lipid moiety to the surface-layer glycoprotein of *Haloferax volcanii*
610 follows protein translocation. Biochem J **366:** 959-964.
611 55. **Feng J, Liu B, Zhang Z, Ren Y, Li Y, Gan F, Huang Y, Chen X, Shen P, Wang L,**
612 **Tang B, Tang XF.** 2012. The complete genome sequence of *Natrinema* sp. J7-2, a
613 haloarchaeon capable of growth on synthetic media without amino acid supplements.
614 PLoS One **7:** e41621.

56. **Mou YZ, Qiu XX, Zhao ML, Cui HL, Oh D, Dyall-Smith ML.** 2012. *Halohasta litorea* gen. nov. sp. nov., and *Halohasta litchfieldiae* sp. nov., isolated from the Daliang aquaculture farm, China and from Deep Lake, Antarctica, respectively. Extremophiles **16:** 895–901.

57. **Honke K, Taniguchi N.** 2002. Sulfotransferases and sulfated oligosaccharides. Med Res Rev **22:** 637-654.

58. **Brissette JL, Weiner L, Ripmaster TL, Model P.** 1991. Characterization and sequence of the *Escherichia coli* stress-induced *psp* operon. J Mol Biol **220:** 35-48.

59. **Bidle KA, Kirkland PA, Nannen JL, Maupin-Furlow JA.** 2008. Proteomic analysis of *Haloferax volcanii* reveals salinity-mediated regulation of the stress response protein PspA. Microbiol **154:** 1436-1443.

60. **Gutierrez C, Devedjian JC.** 1991. Osmotic induction of gene *osmC* expression in *Escherichia coli* K12. J Mol Biol **220:** 959-973.

61. **Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R.** 2005. Genome-wide analysis of the general stress response network in *Escherichia coli*: sigma S-dependent genes, promoters, and sigma factor selectivity. J Bacteriol **187:**1591-1603.

62. **Park SC, Pham BP, Van Duyet L, Jia B, Lee S, Yu R, Han SW, Yang JK, Hahm KS, Cheong GW.** 2008. Structural and functional characterization of osmotically inducible protein C (OsmC) from *Thermococcus kodakaraensis* KOD1. Biochim Biophys Acta **1784**: 783-788.

63. **Esquivel RN, Xu R, Pohlschroder M.** 2013. Novel archaeal adhesion pilins with a conserved N terminus. J Bacteriol **195:** 3808–3818.

64. **Losensky G, Vidakovic L, Klingl A, Pfeifer F, Fröls S.** 2015. Novel pili-like surface structures of *Halobacterium salinarum* strain R1 are crucial for surface adhesion. Front Microbiol **5:** 755.

65. **Jing H, Takagi J, Liu JH, Lindgren S, Zhang RG, Joachimiak A, Wang JH, Springer TA.** 2002. Archaeal surface layer proteins contain beta propeller, PKD, and beta helix domains and are related to metazoan cell surface proteins. Structure **10:**1453-1464.

66. **Makarova KS, Yutin N, Bell SD, Koonin EV.** 2010. Evolution of diverse cell division and vesicle formation systems in Archaea. Nat Rev Microbiol **8:** 731-741.

67. **Confalonieri F, Marsault J, Duguet M.** 1994. SAV, an archaebacterial gene with extensive homology to a family of highly conserved eukaryotic ATPases. J Mol Biol **235:** 396-401.

68. **Maeshima M.** 2000. Vacuolar H^+-pyrophosphatase. Biochim Biophys Acta **1465:** 37–51

69. **Arutyunov D, Arenson B, Manchak J, Frost LS.** 2010. F plasmid TraF and TraH are components of an outer membrane complex involved in conjugation. J Bacteriol **192:** 1730-1734.

Supplementary material figure legends and list of tables

Fig. S1 Number of proteins detected for single filter samples. Metaproteomics were performed on a total of 15 filter samples, representing three distinct size fractions (0.1 – 0.8 μm on 0.1 μm filters; 0.8 – 3 μm on 0.8 μm filters; 3 – 20 μm on 3 μm filters) from 5 distinct depths (0 m, 5 m, 13 m, 24 m, 36 m). Fewer proteins were detected from 0.1 μm filter samples compared to the larger size fraction due to a decrease in the amount of biomass.

Fig. S2 Microscopy of Deep Lake water. The image depicts cells attached to particulate matter from surface water filtered through a 20 μm pre-filter prior to capture on a 3 μm filter. Magnification, 100 x; scale bar, 10 μm.

Fig. S3 Taxonomic composition of the Deep Lake metaproteome. Relative abundance of taxa based on number of identified proteins (blue bars); normalized total spectrum counts (red bars); total number of proteins detected for each taxonomic category (numbers above blue bars).

Fig. S4 Relative abundance of proteins within functional categories. **(A)** *Hht. litchfieldiae*; **(B)** DL31; **(C)** *Hrr. lacusprofundi*. Abundance calculated relative to the number of proteins for the respective organism (blue bars) or relative to the sum of the normalized total spectrum counts (red bars).

Fig. S5 Growth response of *Hht. litchfieldiae* to defined substrates. **(A)** aminoethylphosphonate (AEP); **(B)** dihydroxyacetone (DHA); **(C)** starch.

Table S1 Correlations mentioned in the main text.

Table S2 Pearson correlation analyses performed in R using the normalized total spectrum count across the 15 Deep Lake samples.

Table S3 Complete list of proteins identified in the Deep Lake metaproteome

Table S4 Proteins involved in transport functions from the Deep Lake metaproteome with the best matches to *Hht. litchfieldiae*, DL31 and *Hrr. lacusprofundi*.

689 **Table S5** Proteins involved in carbohydrate uptake and metabolism from the Deep Lake
690 metaproteome with the best matches to *Hht. litchfieldiae*.
691
692 **Table S6** Proteins involved in the uptake and metabolism of nitrogen sources from the Deep
693 Lake metaproteome with the best matches to *Hht. litchfieldiae*, DL31 and *Hrr. lacusprofundi*.
694
695 **Table S7** Proteins involved in motility and taxis from the Deep Lake metaproteome with the
696 best matches to *Hht. litchfieldiae*.
697
698 **Table S8** Proteins implicated in protection against and responses to oxidative stress,
699 photolysis and UV irradiation detected in the Deep Lake metaproteome with the best matches
700 to *Hht. litchfieldiae*, DL31 or *Hrr. lacusprofundi*.
701
702 **Table S9** Proteins assigned to *Hht. litchfieldiae* tADL-II.
703
704 **Table S10** Variant proteins detected for *Hht. litchfieldiae*, tADL, DL31 or *Hrr.*
705 *lacusprofundi*.
706
707 **Table S11** Detected proteins with best matches to *Hht. litchfieldiae,* DL31 or *Hrr.*
708 *lacusprofundi* encoded on contigs with neighbouring genes that best matched to other
709 haloarchaeal species.

Fig. S1

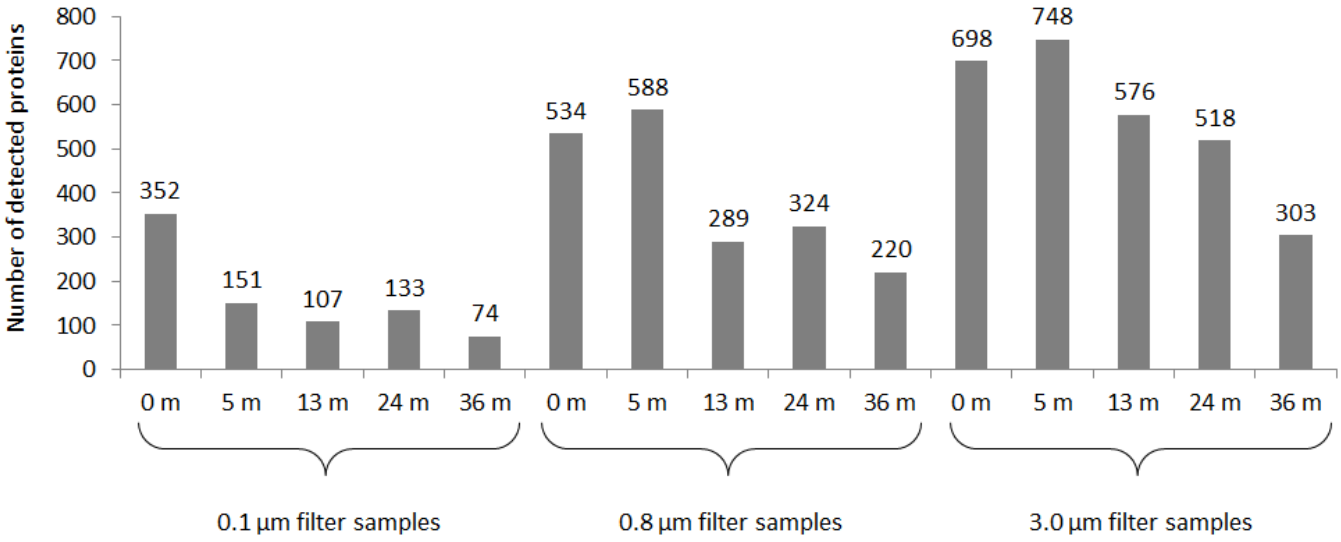

Number of detected proteins
800
700
600
500
400
300
200
100
0
352
151
107
133
74
534
588
289
324
220
698
748
576
518
303
0 m
5 m
13 m
24 m
36 m
0 m
5 m
13 m
24 m
36 m
0 m
5 m
13 m
24 m
36 m
0.1 µm filter samples
0.8 µm filter samples
3.0 µm filter samples

Fig. S1 Number of proteins detected for single filter samples. Metaproteomics were performed on a total of 15 filter samples, representing three distinct size fractions (0.1 – 0.8 μm on 0.1 μm filters; 0.8 – 3 μm on 0.8 μm filters; 3 – 20 μm on 3 μm filters) from 5 distinct depths (0 m, 5 m, 13 m, 24 m, 36 m). Fewer proteins were detected from 0.1 μm filter samples compared to the larger size fraction due to a decrease in the amount of biomass.

Fig. S2

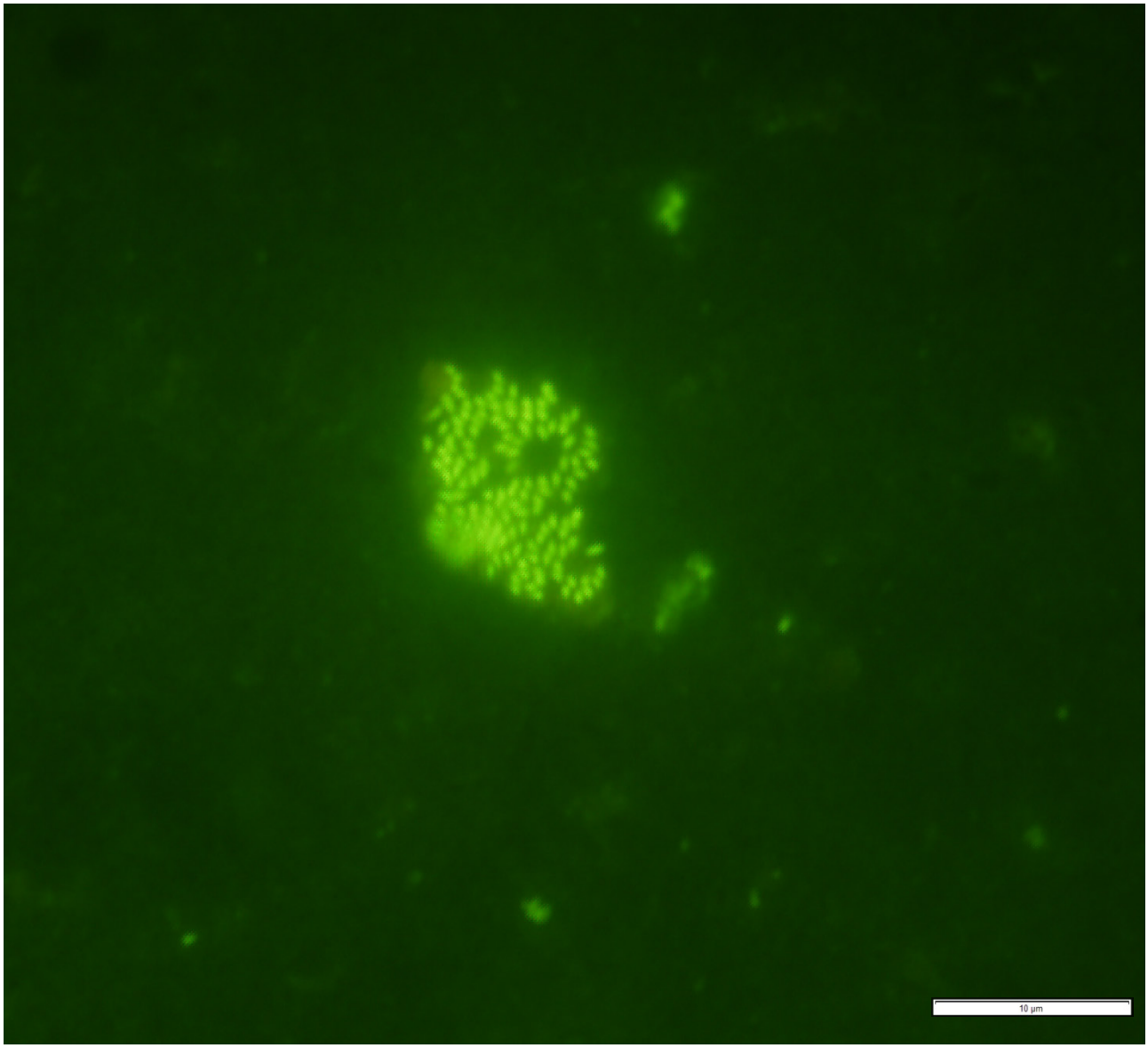


Fig. S2 Microscopy of Deep Lake water. The image depicts cells attached to particulate matter from surface water filtered through a 20 μm pre-filter prior to capture on a 3 μm filter. Magnification, 100 x; scale bar, 10 μm.

Fig. S3

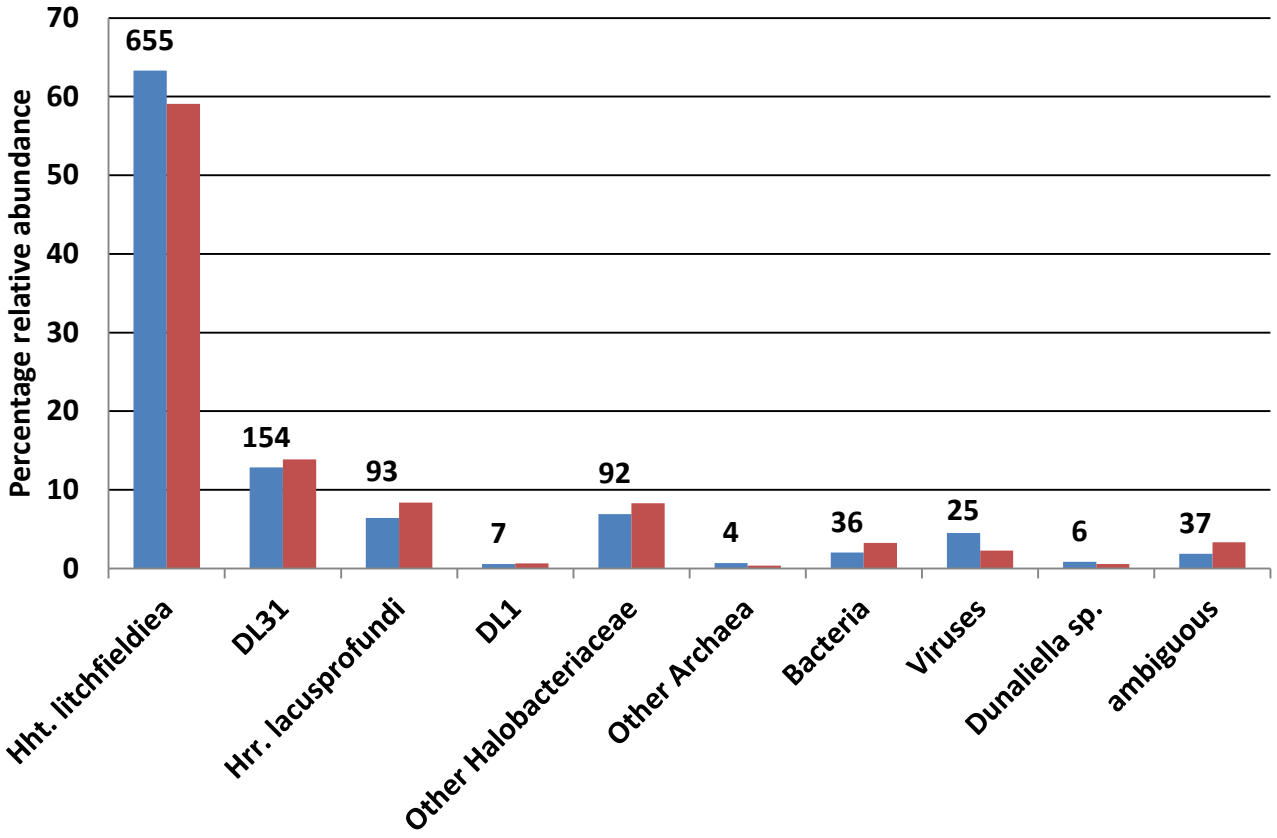
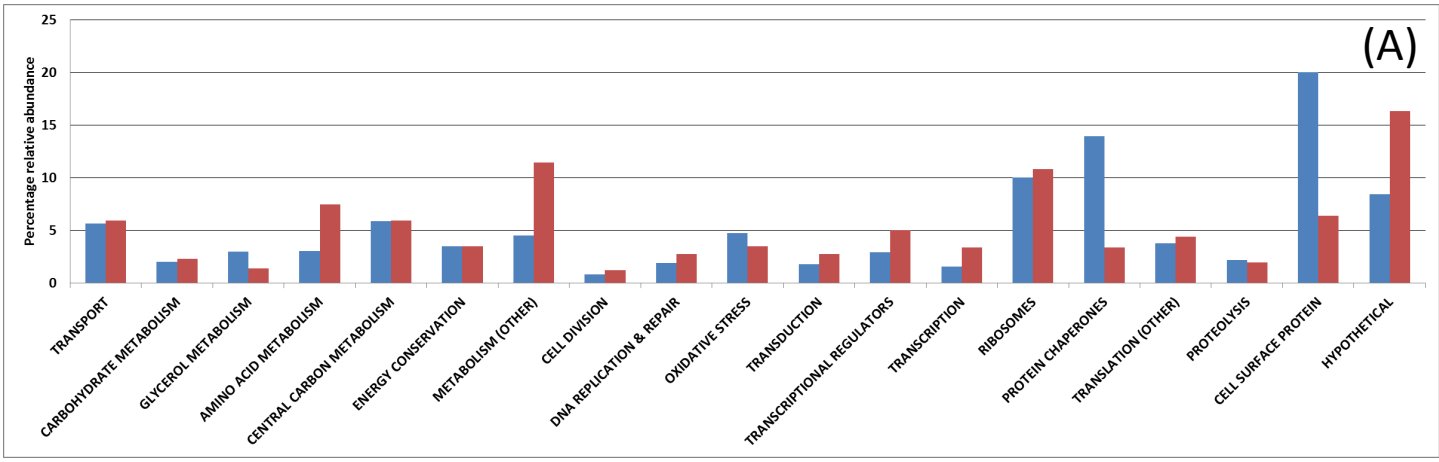
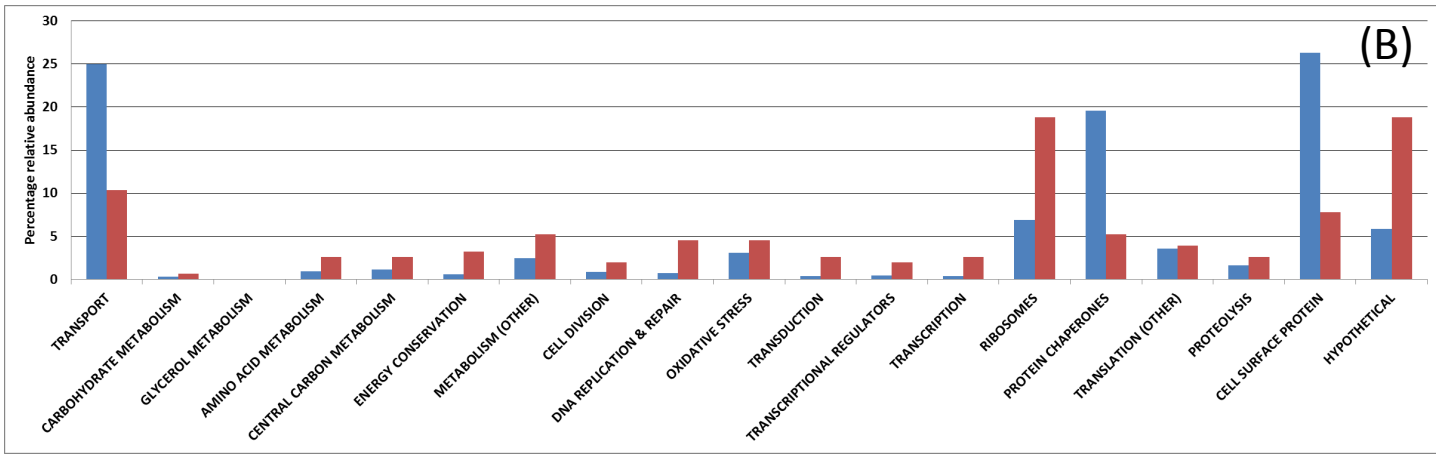

Percentage relative abundance
70
60
50
40
30
20
10
0
655
154
93
7
92
4
36
25
6
37
Hht. litchfieldiea
DL31
Hrr. lacusprofundi
DL1
Other Halobacteriaceae
Other Archaea
Bacteria
Viruses
Dunaliella sp.
ambiguous

Fig. S3 Taxonomic composition of the Deep Lake metaproteome. Relative abundance of taxa based on number of identified proteins (blue bars); normalized total spectrum counts (red bars); total number of proteins detected for each taxonomic category (numbers above blue bars).

Fig. S4


(A)
Percentage relative abundance
25
20
15
10
5
0
TRANSPORT
CARBOHYDRATE METABOLISM
GLYCEROL METABOLISM
AMINO ACID METABOLISM
CENTRAL CARBON METABOLISM
ENERGY CONSERVATION
METABOLISM (OTHER)
CELL DIVISION
DNA REPLICATION & REPAIR
OXIDATIVE STRESS
TRANSDUCTION
TRANSCRIPTIONAL REGULATORS
TRANSCRIPTION
RIBOSOMES
PROTEIN CHAPERONES
TRANSLATION (OTHER)
PROTEOLYSIS
CELL SURFACE PROTEIN
HYPOTHETICAL


(B)
Percentage relative abundance
30
25
20
15
10
5
0
TRANSPORT
CARBOHYDRATE METABOLISM
GLYCEROL METABOLISM
AMINO ACID METABOLISM
CENTRAL CARBON METABOLISM
ENERGY CONSERVATION
METABOLISM (OTHER)
CELL DIVISION
DNA REPLICATION & REPAIR
OXIDATIVE STRESS
TRANSDUCTION
TRANSCRIPTIONAL REGULATORS
TRANSCRIPTION
RIBOSOMES
PROTEIN CHAPERONES
TRANSLATION (OTHER)
PROTEOLYSIS
CELL SURFACE PROTEIN
HYPOTHETICAL

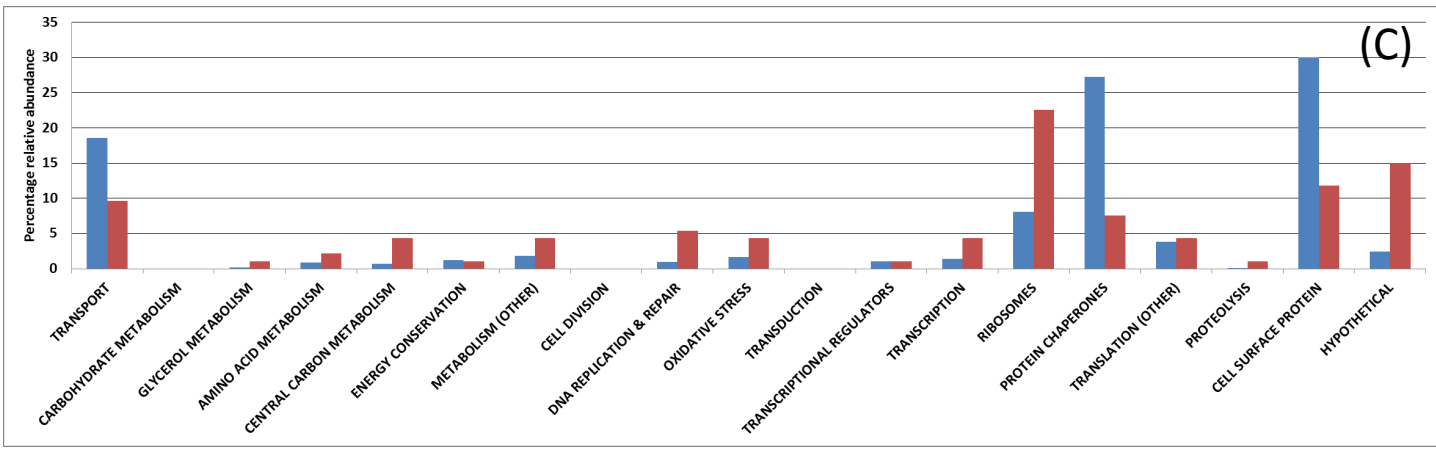
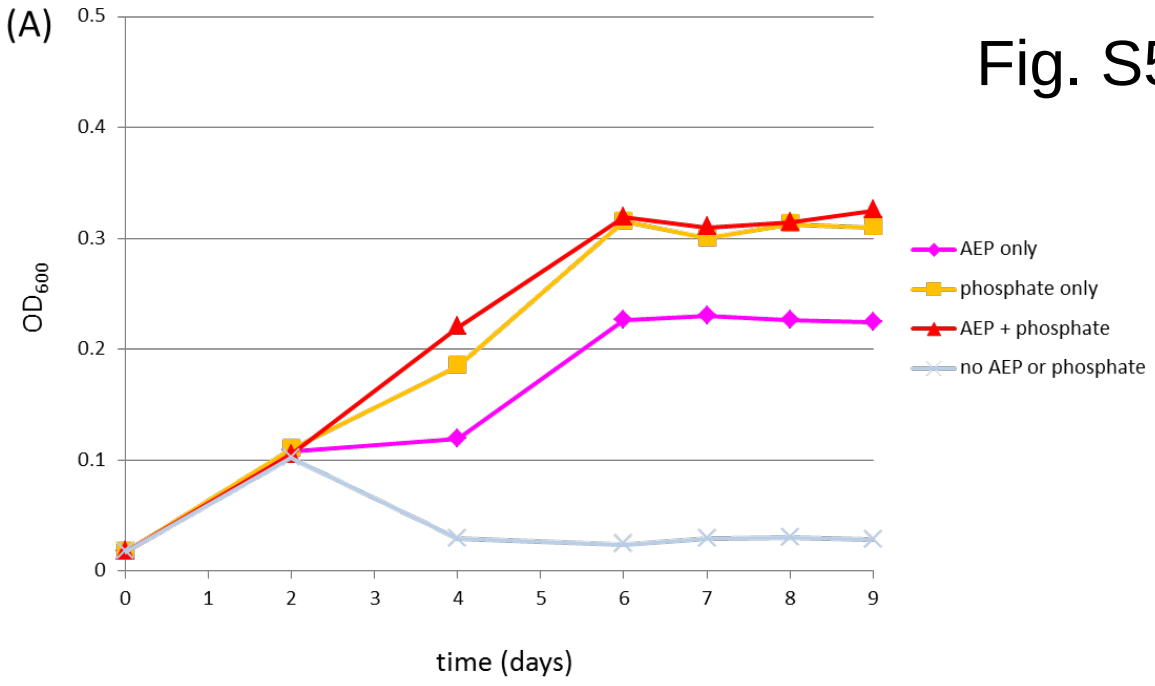
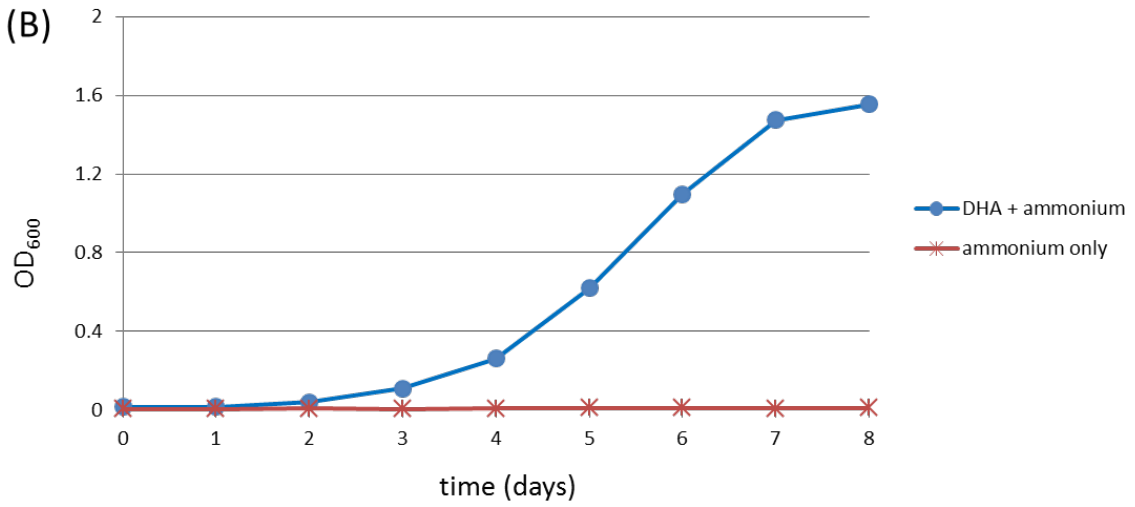

(C)
Percentage relative abundance
35
30
25
20
15
10
5
0
TRANSPORT
CARBOHYDRATE METABOLISM
GLYCEROL METABOLISM
AMINO ACID METABOLISM
CENTRAL CARBON METABOLISM
ENERGY CONSERVATION
METABOLISM (OTHER)
CELL DIVISION
DNA REPLICATION & REPAIR
OXIDATIVE STRESS
TRANSDUCTION
TRANSCRIPTIONAL REGULATORS
TRANSCRIPTION
RIBOSOMES
PROTEIN CHAPERONES
TRANSLATION (OTHER)
PROTEOLYSIS
CELL SURFACE PROTEIN
HYPOTHETICAL

Fig. S4 Relative abundance of proteins within functional categories. **(A)** *Hht. litchfieldiae*; **(B)** DL31; **(C)** *Hrr. lacusprofundi*. Abundance calculated relative to the number of proteins for the respective organism (blue bars) or relative to the sum of the normalized total spectrum counts (red bars).

Fig. S5


(A)
0.5
0.4
0.3
0.2
0.1
0
OD600
0 1 2 3 4 5 6 7 8 9
time (days)
AEP only
phosphate only
AEP + phosphate
no AEP or phosphate


(B)
2
1.6
1.2
0.8
0.4
0
OD600
0 1 2 3 4 5 6 7 8
time (days)
DHA + ammonium
ammonium only

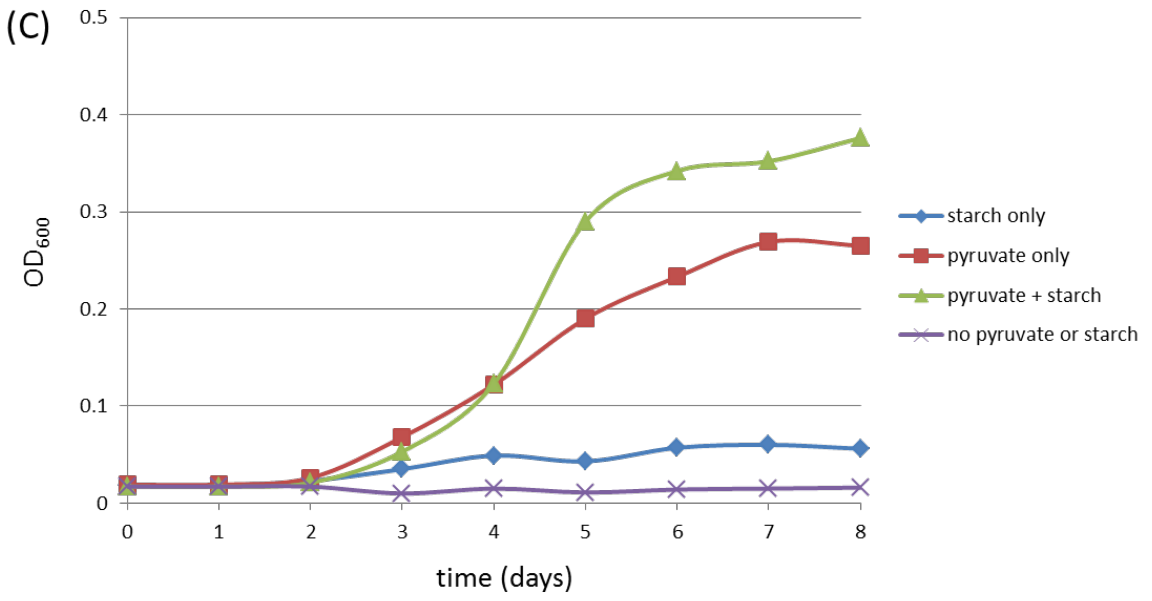

(C)
0.5
0.4
0.3
0.2
0.1
0
OD600
0 1 2 3 4 5 6 7 8
time (days)
starch only
pyruvate only
pyruvate + starch
no pyruvate or starch

Fig. S5 Growth response of *Hht. litchfieldiae* to defined substrates. **(A)** aminoethylphosphonate (AEP); **(B)** dihydroxyacetone (DHA); **(C)** starch.

Table S1 Correlations mentioned in the current study. Fractionation of the Deep Lake biomass according to depth and filter size, comparing the abundances (measured in spectrum counts) of single proteins and proteins from functional categories. Statistically valid positive or negative correlations were determined between spectrum counts for proteins across the 15 filter samples. Pair-wise comparisons of spectrum counts were between individual abundant proteins, functional categories of proteins (the sum of the spectrum count for a functional group, such as tADL ABC transporter proteins) or taxonomic groups of proteins (for example all *Hht. litchfieldiae* tADL proteins vs all *Dunaliella* proteins). Only correlations with a p-value < 0.01 were regarded as statistically significant.. Positive correlations are shown in blue, negative correlations in red.

Correlation	***r***
Species/functional categories	
DL31 transport *vs* *Hrr. lacusprofundi* transport	0.96
Hht. litchfieldiae transport *vs* DL31 transport	-0.91
Hht. litchfieldiae transport *vs* *Hrr. lacusprofundi* transport	-0.89
Species/functional subcategories	
Hht. litchfieldiae archaellins *vs* *Hht. litchfieldiae* central carbon metabolism	-0.95
Hht. litchfieldiae archaellins *vs* *Hht. litchfieldiae* ribosomes	-0.87
DL31 ABC transporter – oligopeptides *vs* DL31 ABC transporter – amino acids	0.94
DL31 ABC transporter – oligopeptides *vs* DL31 ABC transporter – iron	0.89
DL31 ABC transporter – amino acids *vs* DL31 ABC transporter – iron	0.92
DL31 ABC transporters – oligopeptides *vs* *Hrr. lacusprofundi* TRAP transporters	0.76
DL31 ABC transporters – amino acids *vs* *Hrr. lacusprofundi* TRAP transporters	0.80
DL31 ABC transporters – oligopeptides *vs* *Hrr. lacusprofundi* ABC transporters - oligopeptides	0.91
DL31 ABC transporters – oligopeptides *vs* *Hrr. lacusprofundi* ABC transporters – amino acids	0.72

DL31 ABC transporters – amino acids *vs* *Hrr. lacusprofundi* ABC transporters – amino acids	0.72
DL31 ABC transporters – iron *vs* *Hrr. lacusprofundi* ABC transporters – oligopeptides	0.95
DL31 ABC transporters – amino acids *vs* *Hrr. lacusprofundi* ABC transporters – oligopeptides	0.92
DL31 ABC transporters – iron *vs* *Hrr. lacusprofundi* TRAP transporters	0.84
Hrr. lacusprofundi TRAP transporters *vs* *Hrr. lacusprofundi* ABC transporters – oligopeptides	0.90
Individual proteins	
Hht. litchfieldiae archaellin halTADL_1544 (protein #1) *vs* *Hht. litchfieldiae* archaellin halTADL_1812 (protein #3)	0.88
Hht. litchfieldiae archaellin halTADL_1544 (protein #1) *vs* *Hht. litchfieldiae* archaellin halTADL_1813 (protein #7)	0.85
Hht. litchfieldiae archaellin halTADL_1544 (protein #1) *vs* *Hht. litchfieldiae* archaellin halTADL_1810 (protein #21)	0.74
Hht. litchfieldiae archaellin halTADL_1544 (protein #1) *vs* *Hht. litchfieldiae* archaellin halTADL_1811 (protein #80)	0.95
Hht. litchfieldiae archaellin halTADL_1544 (protein #1) *vs* *Hht. litchfieldiae* archaellin halTADL_0078 (protein #95)	0.96
Hht. litchfieldiae archaellin halTADL_1812 (protein #3) *vs* *Hht. litchfieldiae* archaellin halTADL_1813 (protein #7)	0.97
Hht. litchfieldiae archaellin halTADL_1812 (protein #3) *vs* *Hht. litchfieldiae* archaellin halTADL_1813 (protein #12)	0.90
Hht. litchfieldiae archaellin halTADL_1812 (protein #3) *vs* *Hht. litchfieldiae* archaellin halTADL_1810 (protein #21)	0.85
Hht. litchfieldiae archaellin halTADL_1812 (protein #3) *vs* *Hht. litchfieldiae* archaellin halTADL_1811 (protein #80)	0.89
Hht. litchfieldiae archaellin halTADL_1812 (protein #3) *vs* *Hht. litchfieldiae* archaellin halTADL_0078 (protein #95)	0.89
Hht. litchfieldiae archaellin halTADL_1812 (protein #3) *vs* *Hht. litchfieldiae* archaellin halTADL_0078 (protein #118)	0.78
Hht. litchfieldiae archaellin halTADL_1812 (protein #59) *vs* *Hht. litchfieldiae* archaellin halTADL_1810 (protein #21)	0.82

Hht. litchfieldiae archaellin halTADL_1813 (protein #7) *vs* *Hht. litchfieldiae* archaellin halTADL_1813 (protein #12)	0.91
Hht. litchfieldiae archaellin halTADL_1813 (protein #7) *vs* *Hht. litchfieldiae* archaellin halTADL_1810 (protein #21)	0.92
Hht. litchfieldiae archaellin halTADL_1813 (protein #7) *vs* *Hht. litchfieldiae* archaellin halTADL_1811 (protein #80)	0.82
Hht. litchfieldiae archaellin halTADL_1813 (protein #7) *vs* *Hht. litchfieldiae* archaellin halTADL_0078 (protein #95)	0.84
Hht. litchfieldiae archaellin halTADL_1813 (protein #7) *vs* *Hht. litchfieldiae* archaellin halTADL_0078 (protein #118)	0.77
Hht. litchfieldiae archaellin halTADL_1813 (protein #12) *vs* *Hht. litchfieldiae* archaellin halTADL_1810 (protein #21)	0.92
Hht. litchfieldiae archaellin halTADL_1813 (protein #12) *vs* *Hht. litchfieldiae* archaellin halTADL_0078 (protein #118)	0.94
Hht. litchfieldiae archaellin halTADL_1810 (protein #21) *vs* *Hht. litchfieldiae* archaellin halTADL_0078 (protein #118)	0.85
DL31 ABC transporter lipoprotein – oligopeptides Halar_2016 (Protein #9) *vs* DL31 MmpL family membrane transporter Halar_1791 (Protein #29)	0.76
DL31 ABC transporter lipoprotein – iron Halar_0820 (Protein #33) *vs* DL31 MmpL family membrane transporter Halar_1791 (Protein #29)	0.74
Hht. litchfieldiae archaellin halTADL_1544 (protein #1) *vs* *Hht. litchfieldiae* α-amylase halTADL_0142 (Protein #65)	0.86
Hht. litchfieldiae archaellin halTADL_1812 (protein #3) *vs* *Hht. litchfieldiae* α-amylase halTADL_0142 (Protein #65)	0.75
Hht. litchfieldiae archaellin halTADL_1813 (protein #7) *vs* *Hht. litchfieldiae* α-amylase halTADL_0142 (Protein #65)	0.69
Hht. litchfieldiae archaellin halTADL_1811 (protein #80) *vs* *Hht. litchfieldiae* α-amylase halTADL_0142 (Protein #65)	0.86
Hht. litchfieldiae archaellin halTADL_0078 (protein #95) *vs* *Hht. litchfieldiae* α-amylase halTADL_0142 (Protein #65)	0.90
Species	
Hht. litchfieldiae vs Dunaliella	0.68

Table S4 Proteins with transport functions from the Deep Lake metaproteome with the best matches to *Hht. litchfieldiae*, DL31 and *Hrr. lacusprofundi*. Protein numbers are given according to Table S3 (ranked by the sum of the normalized total spectrum count). Sequence identity refers to the amino acid sequence identity of the detected protein to its best match (column denoted "Locus tag") in a BLASTP search. Spectrum count shows the sum of the normalized total spectrum count across all 15 samples. 'nd' denotes 'not determined' due to peptides matching to a protein family (i.e. more than one possible source protein; see Supplementary Information, Materials and Methods). * Denotes a truncated protein.

Protein annotation	Protein #	Locus tag	Sequence identity (%)	Spectrum count
Hht. litchfieldiae				
Phosphate uptake				
phosphate ABC transporter solute-binding protein (PstS)	13	halTADL_2155	99	235
phosphate ABC transporter solute-binding protein (PstS)	61	halTADL_2155	88	92
phosphate ABC transporter solute-binding protein (PstS)	75	halTADL_2155	93	82
phosphate ABC transporter ATPase (PstB)	517	halTADL_2152	100	7.6
phosphate ABC transporter solute-binding protein (PstS)	85	halTADL_1182	51	71
Phosphonate uptake				
phosphonate ABC transporter solute-binding protein (PhnD)	347	halTADL_1334	100	16
Carbohydrate uptake				
carbohydrate ABC transporter solute-binding lipoprotein	265	halTADL_2357	100	25
carbohydrate ABC transporter solute-binding lipoprotein	269	halTADL_2761	100	24
carbohydrate ABC transporter solute-binding lipoprotein	702	halTADL_1911	100	3.2
carbohydrate ABC transporter solute-binding lipoprotein	247	halTADL_2761	84	26
carbohydrate ABC transporter ATPase	1099	halTADL_2764	88	0.4
Amino acid uptake				
branched-chain amino acid ABC transporter solute-binding protein	373	halTADL_2916	100	14
branched-chain amino acid ABC transporter solute-binding protein	202	halTADL_2916	88	33
polar amino acid ABC transporter solute-binding protein	486	halTADL_0024	100	8.5
Ammonium uptake				
ammonium permease (ammonium transporter) (Amt)	220	halTADL_1826	100	30
Urea uptake				
urea ABC transporter solute-binding protein	617	halTADL_0628	100	4.8
Iron uptake				
iron ABC transporter solute-binding protein	159	halTADL_1788	100	40.9
iron ABC transporter solute-binding protein	856	halTADL_1788	83	1.6

Cation transport				
K+ uptake system, TrkA subunit	806	halTADL_3061	100	2.0
K+ uptake system, TrkA subunit	1048	halTADL_3061	89	0.6
K+ uptake system, TrkA subunit	287	halTADL_3258	100	21.2
K+ uptake system, TrkA subunit	854	halTADL_2713	100	1.6
mechanosensitive ion channel (MscS)	436	halTADL_2994	100	10.5
Secretion				
signal peptide peptidase SppA	282	halTADL_2673	100	22.0
PilT protein: Type II/IV secretion system domain + KH domain protein	558	halTADL_0825	100	6.3
SecD/SecF/SecDF export membrane protein	213	halTADL_0787	100	31.4
signal recognition particle Srp54, secretory pathway	500	halTADL_2202	100	7.9
SecD/SecF/SecDF export membrane protein	722	halTADL_0788	100	3.0
TRAP/TTT transport				
Tripartite Tricarboxylate transporter (TTT), solute receptor	138	halTADL_0690	100	45.2
TRAP transporter solute receptor, TAXI family	299	halTADL_0243	100	19.8
Other transport				
thiamine ABC transporter, substrate-binding protein (ThiB)	961	halTADL_2794	100	0.9
nucleoside ABC transporter substrate-binding protein	881	halTADL_2623	nd	1.5
nitrate/sulfonate/bicarbonate ABC transporter solute-binding protein	697	No match (encoded on tADL-II contig)	-	3.3
ABC-type antimicrobial peptide transport system, permease component	492	halTADL_1613	100	8.1
ABC-type antimicrobial peptide transport system, permease component	693	halTADL_1613	88	3.4
heavy metal-exporting ATPase (copper?)	782	halTADL_1767	100	2.3
formate/nitrite transporter	868	halTADL_2501	100	1.5
phosphate/sulfate permease (PiT family)	837	halTADL_3083	100	1.7
RND superfamily family / MMPL (mycobacterial membrane protein large) family protein	957	halTADL_0082	100	0.9
DL31				
Phosphate uptake				
phosphate ABC transporter solute-binding protein (PstS)	533	Halar_1873	100	7.0
Oligopeptide uptake				
oligopeptide/dipeptide ABC transporter solute-binding protein	9	Halar_2016	100	277

oligopeptide/dipeptide ABC transporter solute-binding protein	76	Halar_1439	100	82
oligopeptide/dipeptide ABC transporter solute-binding protein	189	Halar_0722	100	35
oligopeptide/dipeptide ABC transporter solute-binding protein	255	Halar_1146	100	26
oligopeptide/dipeptide ABC transporter solute-binding protein	582	Halar_1285	100	5.5
oligopeptide/dipeptide ABC transporter solute-binding protein	654	Halar_3436	100	4.1
oligopeptide/dipeptide ABC transporter solute-binding protein	818	Halar_2651	100	1.8
oligopeptide/dipeptide ABC transporter solute-binding protein	879	Halar_2024	100	1.5
Amino acid uptake				
branched-chain amino acid ABC transporter solute-binding protein	158	Halar_1569	100	41
branched-chain amino acid ABC transporter solute-binding protein	187	Halar_2890	100	36
branched-chain amino acid ABC transporter solute-binding protein	192	Halar_3433	100	35
Iron uptake				
iron ABC transporter solute-binding protein	33	Halar_0820	100	124
iron ABC transporter solute-binding protein	599	Halar_1080	100	5.1
Other transport				
membrane transporter: MMPL (mycobacterial membrane protein large) family protein / RND superfamily	29	Halar_1791	99	134
uncharacterized transporter (export?), ATPase component	919	Halar_1798	100	1.2
Hrr. lacusprofundi				
Phosphate uptake				
phosphate ABC transporter solute-binding protein (PstS)	326	Hlac_3551	100	17
Oligopeptide uptake				
oligopeptide/dipeptide ABC transporter solute-binding protein	1058	Hlac_0069	100	87
oligopeptide/dipeptide ABC transporter solute-binding protein	630	Hlac_0244	100	4.5
Amino acid uptake				
branched-chain amino acid ABC transporter solute-binding protein*	160	Hlac_2093	100	41
polar amino acid ABC transporter solute-binding protein	891	Hlac_1804	100	1.4
Iron uptake				
iron ABC transporter solute-binding protein	663	Hlac_0162	100	3.9
TRAP/TTT transport				
TRAP transporter solute receptor, TAXI family	56	Hlac_2586	100	97
TRAP transporter solute receptor, TAXI family	144	Hlac_2329	100	44

Other transport				
nucleoside ABC transporter solute-binding protein	496	Hlac_1417	100	8.0

Table S5 Proteins involved in carbohydrate uptake and metabolism from the Deep Lake metaproteome with the best matches to *Hht. litchfieldiae*. Protein numbers are given according to Table S3 (ranked by the sum of the normalized total spectrum count). Sequence identity refers to the amino acid sequence identity of the detected protein to its best match (column denoted "Locus tag") in a BLASTP search. Spectrum count shows the sum of the normalized total spectrum count across all 15 samples. *Denotes a match to a protein that is represented by an incomplete (truncated) gene on contig.

Protein annotation	Protein #	Locus tag	Sequence identity (%)	Spectrum count
Carbohydrate uptake				
carbohydrate ABC transporter solute-binding lipoprotein	265	halTADL_2357	100	25
carbohydrate ABC transporter solute-binding lipoprotein	269	halTADL_2761	100	24
carbohydrate ABC transporter solute-binding lipoprotein	702	halTADL_1911	100	3.2
carbohydrate ABC transporter solute-binding lipoprotein	247	halTADL_2761	84	26
carbohydrate ABC transporter ATPase	1099	halTADL_2764	88	0.4
Polysaccharide/starch degradation				
α-amylase (glycosyl hydrolase, family 13)	65	halTADL_0142	100	90
α-amylase (glycosyl hydrolase, family 13)	193	halTADL_0142	84	35
α-amylase (glycosyl hydrolase, family 13)	543	halTADL_0142	94	6.8
glucan 1,4-α-glucosidase (glucoamylase) (glycosyl hydrolase, family 15)	557	halTADL_0141	100	6.4
4-α-glucanotransferase (amylomaltase) (glycosyl hydrolase, family 77)	710	halTADL_2529	100	3.1
Glycerol catabolism				
glycerol kinase (GlpK)	17	halTADL_2249	100	206
glycerol kinase (GlpK)	28	halTADL_2249	96	13
glycerol kinase (GlpK)	459	halTADL_0681	100	135
glycerol kinase (GlpK)*	381	halTADL_0681	100	9.5
glycerol-3-phosphate dehydrogenase (GlpA)	195	halTADL_2244	100	34
glycerol-3-phosphate dehydrogenase (GlpA)	327	halTADL_2244	87	17
dihydroxyacetone (DHA) kinase, L subunit (DhaL)	266	halTADL_2259	100	25
dihydroxyacetone (DHA) kinase, L subunit (DhaL)	982	halTADL_2259	92	0.9
dihydroxyacetone (DHA) kinase, K subunit (DhaK)	140	halTADL_2260	94	45
Glycosylation / Capsular polysaccharide				
glucose-1-phosphate thymidylyltransferase (RfbA, RffH)	37	halTADL_3353	100	118
glucose-1-phosphate thymidylyltransferase (RfbA, RffH)	421	halTADL_3353	93	11

glycosyl transferase group 1 (possible α--D-glucan synthase)	780	halTADL_2565	100	2.3
nucleoside-diphosphate-sugar epimerase	497	halTADL_3057	100	8.0
NUDIX hydrolase (NUDIX = NUcleoside DIphosphate linked to some other moiety X)	1097	halTADL_2550	100	0.4
NUDIX hydrolase (NUDIX = NUcleoside DIphosphate linked to some other moiety X)	1053	halTADL_3253	100	0.6
oligosaccharyltransferase AglB	540	halTADL_2411	100	6.8
Carbohydrate metabolism (other)				
carbohydrate kinase, FGGY (possible xylulokinase [XylB])	1041	halTADL_2660	100	0.6
phosphoglucomutase/phosphomannomutase	181	halTADL_1712	100	36
ribose 5-phosphate isomerase A (RpiA)	1012	halTADL_1707	100	0.6
Emden-Meyerhof (EM) pathway				
phosphoglucose isomerase	528	halTADL_0801	100	7.2
2-dehydro-3-deoxy-D-gluconate (KDG) kinase (ribokinase family) (KdgK)	1038	halTADL_2089	100	0.6
fructose-1,6-bisphosphate aldolase, class I (FbaB)	171	halTADL_0575	100	39
fructose-1,6-bisphosphate aldolase, class I (FbaB)	733	halTADL_0575	92	2.8
fructose 1,6-bisphosphate aldolase (multifunctional)	225	halTADL_3234	100	29
fructose 1,6-bisphosphate aldolase (multifunctional)	420	halTADL_3234	96	11
fructose-1,6-bisphosphate aldolase, class II (FbaA)	845	halTADL_3223	100	1.6
triosephosphate isomerase (TpiA)	175	halTADL_2532	100	37
triosephosphate isomerase (TpiA)	249	halTADL_2532	89	26
Entner-Doudoroff (ED) Pathway (semiphosphorylative)				
gluconate dehydratase (GnaD)	216	halTADL_0374	100	31
2-dehydro-3-deoxy-D-gluconate (KDG) kinase (ribokinase family) (KdgK)	1038	halTADL_2089	100	0.6
2-dehydro-3-deoxyphosphogluconate (KDPG) aldolase (Eda)	355	halTADL_0882	100	15
ED/EM common pathway				
glyceraldehyde-3-phosphate dehydrogenase (NAD(P)+-dependent), type I (Gap)	110	halTADL_0817	100	54
glyceraldehyde-3-phosphate dehydrogenase (NAD(P)+-dependent), type I (Gap)	591	halTADL_0817	90	5.2
phosphoglycerate kinase (Pgk)	272	halTADL_0816	100	24
phosphoglycerate kinase (Pgk)	833	halTADL_0816	94	1.7
enolase (phosphopyruvate hydratase) (Eno)	25	halTADL_2780	100	142
enolase (phosphopyruvate hydratase) (Eno)	391	halTADL_2780	90	12
pyruvate kinase (Pyk)	334	halTADL_3014	100	17

pyruvate kinase (Pyk)	695	halTADL_3014	91	3.4
Gluconeogenesis				
phosphoenolpyruvate (PEP) synthase (Pps)	451	halTADL_1011	79	9.8

Table S6 Proteins involved in involved in the uptake and metabolism of nitrogen sources from the Deep Lake metaproteome with the best matches to *Hht. litchfieldiae*, DL31, and *Hrr. lacusprofundi*. Protein numbers are given according to Table S3 (ranked by the sum of the normalized total spectrum count). Sequence identity refers to the amino acid sequence identity of the detected protein to its best match (column denoted "Locus tag") in a BLASTP search. Spectrum count shows the sum of the normalized total spectrum count across all 15 samples. 'nd' denotes 'not determined' due to peptides matching to a protein family (i.e. more than one possible source protein; see Supplementary Information, Materials and Methods).

Protein annotation	Protein #	Locus tag	Sequence identity (%)	Spectrum count
Hht. litchfieldiae				
Protein/peptide digestion (extracytoplasmic)				
halolysin (peptidase S8 and S53 subtilisin kexin sedolisin)	477	halTADL_1514	100	8.7
aminopeptidase (peptidase family M42)	723	halTADL_0101	nd	3.0
Amino acid uptake				
branched-chain amino acid ABC transporter solute-binding protein	373	halTADL_2916	100	14
branched-chain amino acid ABC transporter solute-binding protein	202	halTADL_2916	88	33
polar amino acid ABC transporter solute-binding protein	486	halTADL_0024	100	8.5
Ammonium uptake & assimilation				
ammonium permease (ammonium transporter) (Amt)	220	halTADL_1826	100	30
glutamine synthetase (GlnA) → Gln	88	halTADL_3423	100	67
glutamate synthase (GltB) → Glu	760	halTADL_0125	100	2.5
Amino acid biosynthesis				
aspartate aminotransferase (AspB) → Asp, Phe	1086	halTADL_0403	100	0.4
aspartate aminotransferase (AspB) → Asp, Phe	173	halTADL_3081	100	38
aspartate aminotransferase (AspB) → Asp, Phe	619	halTADL_3081	97	4.8
carbamoyl-phosphate synthase, large subunit (CarB) → Arg	469	halTADL_0988	100	9.1
aspartate kinase (LysC) → Lys, Thr, Met	527	halTADL_1916	100	7.3
aspartate kinase (LysC) → Lys, Thr, Met	905	halTADL_1916	90	1.3
aspartate-semialdehyde dehydrogenase (Asd) → Lys, Thr, Met	642	halTADL_0714	100	4.3
homoserine dehydrogenase (MetL) → Lys, Thr, Met	430	halTADL_0649	100	11
threonine synthase (ThrC) → Thr	655	halTADL_2266	97	4.0
cystathionine gamma-synthase (MetB) or O-acetylhomoserine (thiol)-lyase (MetY) → Met	553	halTADL_1890	100	6.5

5-methyltetrahydropteroyltriglutamate -homocysteine methyltransferase (MetE) → Met	487	halTADL_0179	100	8.4
2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase (DapD) → Lys	402	halTADL_0281	nd	12
glutamate-5-semialdehyde dehydrogenase (ProA) → Pro	600	halTADL_2358	nd	5.1
pyrroline-5-carboxylate reductase (ProC) → Pro	588	halTADL_2360	100	5.3
ATP phosphoribosyltransferase (HisG) → His	602	halTADL_1729	100	5.1
phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (HisA) → His	787	halTADL_1799	100	2.2
imidazoleglycerol-phosphate dehydratase (HisB) → His	442	halTADL_1797	100	10
phosphoserine phosphatase (SerB) → Ser	590	halTADL_1053	100	5.3
phosphoserine phosphatase (SerB) → Ser	309	halTADL_2046	96	19
D-3-phosphoglycerate dehydrogenase (Ser A) → Ser	131	halTADL_2045	100	47
D-3-phosphoglycerate dehydrogenase (Ser A) → Ser	236	halTADL_2045	93	28
D-3-phosphoglycerate dehydrogenase (Ser A) → Ser	626	halTADL_0712	100	4.6
glycine hydroxymethyltransferase (GlyA) → Gly	501	halTADL_3114	nd	7.9
rhodanese-like protein / thiosulfate sulfurtransferase → Cys?	83	halTADL_2750	100	74
rhodanese-like protein / thiosulfate sulfurtransferase → Cys?	258	halTADL_2750	92	25
fructose 1,6-bisphosphate aldolase (multifunctional) → Trp, Phe, Tyr	225	halTADL_3234	100	29
fructose 1,6-bisphosphate aldolase (multifunctional) → Trp, Phe, Tyr	420	halTADL_3234	96	11
2-amino-3,7-dideoxy-D-threo-hept-6-ulosonate synthase → Trp, Phe, Tyr	171	halTADL_0575	100	39
2-amino-3,7-dideoxy-D-threo-hept-6-ulosonate synthase → Trp, Phe, Tyr	733	halTADL_0575	92	2.8
dehydroquinate synthase II → Trp, Phe, Tyr	408	halTADL_0574	100	12
shikimate kinase (AroB) → Trp, Phe, Tyr	984	halTADL_2582	100	0.9
tryptophan synthase, alpha subunit (TrpA) → Trp	888	halTADL_0576	nd	1.5
anthranilate phosphoribosyltransferase (TrpD) → Trp	390	halTADL_0889	100	13
anthranilate phosphoribosyltransferase (TrpD) → Trp	290	halTADL_3066	100	21
anthranilate phosphoribosyltransferase (TrpD) → Trp	1049	halTADL_3066	92	0.6
prephenate dehydratase (PheA2) → Phe	701	halTADL_2073	100	3.3
branched-chain amino acid aminotransferase (IlvE) → Leu, Val, Ile	841	halTADL_1961	nd	1.6
3-isopropylmalate dehydrogenase (LeuB) → Leu	385	halTADL_0366	100	13
3-isopropylmalate dehydrogenase (LeuB) → Leu	1084	halTADL_0366	93	0.4

3-isopropylmalate/(R)-2-methylmalate dehydratase, large subunit (LeuC) → Leu, Ile	890	halTADL_0364	nd	1.4
3-isopropylmalate/(R)-2-methylmalate dehydratase, small subunit (LeuD) → Leu, Ile	894	halTADL_0365	nd	1.4
acetolactate synthase, small subunit (IlvH) → Ile, Val	560	halTADL_0361	100	6.3
ketol-acid reductoisomerase (IlvC) →Ile, Val	251	halTADL_0362	100	26
dihydroxy-acid dehydratase (IlvD) → Ile, Val	323	halTADL_2417	nd	18
2-isopropylmalate synthase (LeuA) → Leu	729	halTADL_0359	100	2.9
citramalate synthase (CimA) → Ile	605	halTADL_1156	100	5.0
Amino acid degradation				
glutamate dehydrogenase (GdhA) ← Glu	1013	halTADL_1757	81	0.6
S-adenosylmethionine synthetase (Mat) ← Met	458	halTADL_3028	100	9.5
S-adenosylhomocysteine hydrolase (AchY) ← Met	826	halTADL_1723	100	1.8
2-oxoacid dehydrogenase complex, E2 component ← Leu, Val, Ile	631	halTADL_2147	100	4.5
2-oxoacid dehydrogenase complex, dihydrolipoamide dehydrogenase ← Leu, Val, Ile	679	halTADL_2144	100	3.6
Urea uptake & degradation				
urea ABC transporter solute-binding protein	617	halTADL_0628	100	4.8
urease, beta subunit (UreB)	1089	halTADL_0634	nd	0.4
DL31				
Protein/peptide digestion (extracytoplasmic)				
halolysin (peptidase S8 and S53 subtilisin kexin sedolisin)	1080	Halar_3678	100	0.4
aminopeptidase (peptidase family M42)	745	Halar_3640	100	2.7
Amino acid uptake				
branched-chain amino acid ABC transporter solute-binding protein	158	Halar_1569	100	41.1
branched-chain amino acid ABC transporter solute-binding protein	187	Halar_2890	100	35.5
branched-chain amino acid ABC transporter solute-binding protein	192	Halar_3433	100	35.2
Oligopeptide/dipeptide uptake				
oligopeptide/dipeptide ABC transporter solute-binding protein	9	Halar_2016	100	277.4
oligopeptide/dipeptide ABC transporter solute-binding protein	76	Halar_1439	100	81.8
oligopeptide/dipeptide ABC transporter solute-binding protein	189	Halar_0722	100	35.4
oligopeptide/dipeptide ABC transporter solute-binding protein	255	Halar_1146	100	25.7

oligopeptide/dipeptide ABC transporter solute-binding protein	582	Halar_1285	100	5.5
oligopeptide/dipeptide ABC transporter solute-binding protein	654	Halar_3436	100	4.1
oligopeptide/dipeptide ABC transporter solute-binding protein	818	Halar_2651	100	1.8
oligopeptide/dipeptide ABC transporter solute-binding protein	879	Halar_2024	100	1.5
Amino acid biosynthesis & degradation				
branched chain amino acid aminotransferase (IlvE) → Leu, Val, Ile	371	Halar_2889	100	13.7
pyrroline-5-carboxylate dehydrogenase (RocA) → Pro	743	Halar_1051	100	2.7
3-isopropylmalate dehydrogenase (LeuB) → Leu	1026	Halar_2164	100	0.6
glutamate dehydrogenase (GdhA) ← Glu	378	Halar_0758	100	13.3
Hrr. lacusprofundi				
Amino acid uptake				
branched-chain amino acid ABC transporter solute-binding protein (gene appears to be interrupted by a transposase)	160	Hlac_2093	100	40.5
polar amino acid ABC transporter solute-binding protein	891	Hlac_1804	100	1.4
Oligopeptide/dipeptide uptake				
oligopeptide/dipeptide ABC transporter solute-binding protein	71	Hlac_0069	100	86.5
oligopeptide/dipeptide ABC transporter solute-binding protein	630	Hlac_0244	100	4.5
Ammonium uptake & assimilation				
glutamine synthetase (GS) (GlnA) → Gln	375	Hlac_2374	100	13.5
Amino acid biosynthesis				
histidinol-phosphate aminotransferase (HisC) → His	1017	Hlac_0235	100	0.6

Table S7 Proteins involved in involved in motility, taxis and adhesion from the Deep Lake metaproteome with the best matches to *Hht. litchfieldiae*. Protein numbers are given according to Table S3 (ranked by the sum of the normalized total spectrum count). Sequence identity refers to the amino acid sequence identity of the detected protein to its best match (column denoted "Locus tag") in a BLASTP search. Spectrum count shows the sum of the normalized total spectrum count across all 15 samples.

Protein annotation	Protein #	Locus tag	Sequence identity (%)	Spectrum count
Archaella				
archaellin (FlaA or FlaB)	1	halTADL_1544	100	695
archaellin (FlaA or FlaB)	786	halTADL_1544	75	2.3
archaellin (FlaA or FlaB)	3	halTADL_1812	100	497
archaellin (FlaA or FlaB)	59	halTADL_1812	77	94
archaellin (FlaA or FlaB)	7	halTADL_1813	100	351
archaellin (FlaA or FlaB)	12	halTADL_1813	76	239
archaellin (FlaA or FlaB)	21	halTADL_1810	100	169
archaellin (FlaA or FlaB)	80	halTADL_1811	100	79
archaellin (FlaA or FlaB)	118	halTADL_0078	100	52
archaellin (FlaA or FlaB)	95	halTADL_0078	75	59
archaellar protein FlaG	429	halTADL_1803	100	12
archaellar protein FlaC or FlacD or FlacE	761	halTADL_1805	100	2.5
Taxis				
bacteriorhodopsin	752	halTADL_1952	100	2.6
methyl-accepting chemotaxis sensory transducer (HtrII) (for sensory rhodopsin II)	245	halTADL_3325	100	26
globin domain + methyl-accepting chemotaxis sensory transducer	235	halTADL_0074	100	28
heme-based aerotactic transducer HemAT	351	halTADL_1627	100	15
PBS_HEAT protein; taxis signaling	684	halTADL_1768	100	3.4
methyl-accepting chemotaxis sensory transducer with Pas/Pac sensor	1093	halTADL_1218	100	0.4
signal transduction protein with CBS domains	292	halTADL_1865	100	21
chemotaxis signal transduction protein CheW	473	halTADL_1838	100	8.9
chemotaxis signal transduction protein CheW	925	halTADL_1838	91	1.2
chemotaxis response regulator CheY	300	halTADL_1808	100	20

chemotaxis response regulator CheY	566	halTADL_1808	94	6.0
response regulator receiver protein	538	halTADL_2200	100	6.9
response regulator receiver protein	1036	halTADL_1816	100	0.6
response regulator receiver domain + HalX domain	102	halTADL_0055	96	29
KaiC domain	278	halTADL_1815	100	23
KaiC domain	924	halTADL_1815	94	1.2

Table S8 Proteins implicated in protection against and responses to oxidative stress, photolysis and UV irradiation detected in the Deep Lake metaproteome for the haloarchaeal species *Hht. litchfieldiae*, DL31 or *Hrr. lacusprofundi*.

Proteins	Locus tags	Specific function
Oxidative stress		
manganese superoxide dismutase (SOD)	halTADL_2687, Halar_1640, Hlac_2515	Degradation of harmful oxygen radicals.
methionine sulfoxide reductase	halTADL_1172	Reduction of methionine sulfoxide (the oxidized form of methionine, caused by ROS) back to methionine, to reactivate damaged proteins.
thioredoxin	halTADL_1187/1756/2077/2563, Halar_3305, Hlac_0372	Protection of sulfhydryl groups of cysteine residues from forming disulfide linkages under conditions of oxidative stress. Thioredoxin also acts as an electron donor for enzymes such as methionine sulfoxide reductase and peroxiredoxin (Meyer *et al.,* 2009).
glutaredoxin	halTADL_0399/2104	
peroxiredoxin	halTADL_0067; Halar_1849	Protective antioxidant role via their peroxidase activity against hydrogen peroxide and organic hydroperoxides (Wood *et al.,* 2003).
Dps ferritins (miniferritins)	halTADL_1068, Halar_0843/0845, Hlac_0536	Sequester intracellular iron; these are particularly important when protein turnover releases iron from degraded proteins (Theil *et al.,* 2007). Fe^{2+} reacts with hydrogen peroxide to generate hydroxyl radicals that can damage cellular components, and with O_2 to generate Fe^{3+} which precipitates ('rusts') at physiological pH.
universal stress protein (UspA)	halTADL_0697/1044/1904/2110/ 2112/2276/2351	Inferred to protect cells against oxidative damage to DNA (Kvint *et al.,* 2003; Nachin *et al.,* 2005).
RosR transcriptional regulator	halTADL_0352/1645; Halar_0879	Regulates gene expression in response to oxidative stress in, including upregulation of SOD (Sharma *et al.,* 2012).
PitA (haloarchaeal hypothetical protein)	halTADL_1349	PitA proteins are so far unique to haloarchaea; although of unknown function, they are inferred to have a function associated with life in hypersaline environments, particularly exposure to limited oxygen availability and ROS neutralization (Bab-Dinitz *et al.,* 2006; Martinez-Espinosa *et al.,* 2015).
Photolysis		
dodecin	halTADL_3198, Halar_2184	Small dodecameric flavoprotein that binds riboflavin and protects it from light-induced degradation (Grininger *et al.,* 2009). Photolytic degradation of riboflavin also leads to toxic derivatives such as lumichrome, which is involved in the generation of cytotoxic singlet oxygen (O_2*) via transfer of light energy to O_2 (Sikorski *et al.,* 2001; Grininger *et al.,* 2009).

UV irradiation		
RadA (Rad51/RecA recombinase homolog)	halTADL_1827/2135, Halar_3361, Hlac_2624	Catalyzes strand invasion and exchange during homologous recombination, and appears to be critical to the haloarchaeal UV response by permitting rescue of stalled replication forks and/or facilitate recombinational repair (Woods and Dyall-Smith, 1997; McCready *et al.,* 2005; Boubriak *et al.,* 2008).
RecJ-like exonuclease	halTADL_1000	Involved in processing stalled forks in UV-damaged DNA, possibly by making DNA lesions at stalled forks accessible for repair (Boubriak *et al.,* 2008).
topoisomerase VI	halTADL_3021 (subunit B [Top6B])	Involved in processing stalled forks in UV-damaged DNA, possibly by ATP-dependent nicking-closing activity as well as ability to generate double-strand breaks (Boubriak *et al.,* 2008).
ribonucleotide reductase (RNR; adenosylcobalamin-dependent)	halTADL0884	Catalyzes the rate-limiting step in DNA synthesis from RNA; inferred to have a function in excision repair (McCready *et al.,* 2005).
UvrD helicase domain protein	halTADL_2299	Possibly involved in post-incision events of nucleotide excision repair, as for bacterial UvrD helicase.
replication protein A (RPA) ssDNA-binding complex components	halTADL_3434/_2569 (RPA32), halTADL_3433 (RPA41), Halar_3463 (RPA32)	RPA-ssDNA complexes are proposed to form during DNA-damage repair pathways in haloarchaea, as for eukaryotic RPA-ssDNA complexes (McCready *et al.,* 2005).

Table S9 Proteins assigned to *Hht. litchfieldiae* tADL-II. Protein numbers are given according to Table S3 (ranked by the sum of the normalized total spectrum count). Sequence identity refers to the amino acid sequence identity of the detected protein to its best match (column denoted "Locus tag") in a BLASTP search. Spectrum count shows the sum of the normalized total spectrum count across all 15 samples. Highlighted in purple are tADL-II proteins with a higher spectrum count than the respective tADL protein.

Protein annotation	Protein #	Matching tADL locus tag	Sequence identity (%)	Spectrum count
Amino acid metabolism				
3-isopropylmalate dehydrogenase (LeuB)	1084	halTADL_0366	93	0.4
agmatinase (SpeB)	374	halTADL_1131	95	13
glutamate dehydrogenase (GdhA)	1013	halTADL_1757	81	0.6
aspartate kinase (LysC)	905	halTADL_1916	90	1.3
D-3-phosphoglycerate dehydrogenase (Ser A)	236	halTADL_2045	93	28
phosphoserine phosphatase (SerB)	309	halTADL_2046	96	19
anthranilate phosphoribosyltransferase (TrpD)	1049	halTADL_3066	92	0.6
aspartate aminotransferase (AspB)	619	halTADL_3081	97	4.8
glutamine synthetase (GS) (GlnA)	319	halTADL_3423	97	18
Carbohydrate metabolism				
alpha-amylase (glycosyl hydrolase, family 13)	193	halTADL_0142	84	35
glucose-1-phosphate thymidylyltransferase (RfbA, RffH)	421	halTADL_3353	93	11
Cell division				
cell division protein FtsA	467	halTADL_0130	96	9.2
VCP-like protein (2 x CDC48 domains + 2 x AAA family ATPase domains)	927	halTADL_2740	95	1.2
Cell surface				
invasin/intimin cell-adhesion domain protein (Sec signal)	577	No match to tADL – unique to tADL-II	-	5.5
archaellin FlaA or FlaB	95	halTADL_0078	75	59
hypothetical protein (TAT signal)	490	halTADL_0878	50	8.3
adhesion pilin (PilA)	666	halTADL_1387	66	3.9
hypothetical protein (Sec signal, PGF-CTERM, C-terminal transmembrane helix)	126	halTADL_1403	29	49
archaellin FlaA or FlaB	786	halTADL_1544	75	2.3

hypothetical protein (TAT signal)	370	halTADL_1761	40	14
hypothetical protein (Sec signal, PGF-CTERM, C-terminal transmembrane helix)	361	halTADL_1765	76	14
archaellin FlaA or FlaB	59	halTADL_1812	77	94
archaellin FlaA or FlaB	12	halTADL_1813	76	239
Central carbon metabolism				
pyruvate:ferredoxin oxidoreductase, alpha subunit (PorA)	691	halTADL_0382	93	3.4
phosphoenolpyruvate carboxylase (Ppc)	572	halTADL_0401	94	5.9
fructose-1,6-bisphosphate aldolase, class I (FbaB)	733	halTADL_0575	92	2.8
phosphoglycerate kinase (Pgk)	833	halTADL_0816	94	1.7
glyceraldehyde-3-phosphate dehydrogenase (NAD(P)+-dependent), type I (Gap)	591	halTADL_0817	90	5.2
phosphoenolpyruvate (PEP) synthase (Pps)	451	halTADL_1011	79	9.8
acetate : CoA ligase (Acs)	449	halTADL_1017	92	9.8
triosephosphate isomerase (TpiA)	249	halTADL_2532	89	26
enolase (phosphopyruvate hydratase) (Eno)	391	halTADL_2780	90	12
aconitate hydratase (AcnA)	401	halTADL_2902	96	12
pyruvate kinase (Pyk)	695	halTADL_3014	91	3.4
fructose 1,6-bisphosphate aldolase (multifunctional)	420	halTADL_3234	96	11
DNA replication & repair				
methylated-DNA-[protein]-cysteine S-methyltransferase (Ogt)	1031	halTADL_0579	82	0.6
DNA polymerase sliding clamp subunit (PCNA homolog)	244	halTADL_1713	97	27
DNA repair and recombination protein RadA	284	halTADL_2135	90	22
Energy conservation				
cytochrome c oxidase, subunit II (CoxB)	959	halTADL_1060	82	0.9
inorganic pyrophosphatase (Ppa)	937	halTADL_1644	88	1.0
ATP synthase, K subunit (AtpK)	293	halTADL_1940	80	21
ATP synthase, beta subunit (AtpB)	406	halTADL_1945	92	12
ferredoxin	42	halTADL_2137	96	109
NADPH-dependent F420 reductase	1096	halTADL_2320	91	0.4
Glycerol metabolism				
glycerol-3-phosphate dehydrogenase (GlpA)	327	halTADL_2244	87	17
glycerol kinase (GlpK)	381	halTADL_2249	96	13
dihydroxyacetone (DHA) kinase, L subunit (DhaL)	982	halTADL_2259	92	0.9

dihydroxyacetone (DHA) kinase, K subunit (DhaK)	140	halTADL_2260	94	45
Hypothetical				
hypothetical protein	431	halTADL_0015	94	10
nucleic acid-binding/OB-fold/TRAM domain protein	267	halTADL_0109	93	24
DUF964 / YheA/YmcA domain	777	halTADL_0133	88	2.4
DUF655 (predicted RNA-binding domain)	428	halTADL_0183	89	10
hypothetical protein	887	halTADL_0395	85	1.5
predicted RNA-binding protein containing KH domain, possibly ribosomal protein	1088	halTADL_0545	81	0.4
DUF827 / WEB family domain	183	halTADL_0555	77	36
selT/selW/selH selenoprotein domain / Rdx domain	979	halTADL_1063	81	0.9
ThiJ/PfpI domain-containing protein	199	halTADL_1769	83	33
hypothetical protein (DUF4382)	215	halTADL_2062	87	31
hypothetical protein	307	halTADL_2560	89	19
hypothetical protein	629	halTADL_2576	95	4.5
hypothetical protein	1100	halTADL_3036	79	0.4
nucleic acid binding OB-fold tRNA/helicase-type	150	halTADL_3218	86	42
DUF4013 (4 x transmembrane domains)	727	halTADL_3238	73	3.0
Metabolism (other)				
nucleoside-diphosphate kinase (Ndk)	641	halTADL_0169	84	4.4
orotate phosphoribosyltransferase (PyrE)	749	halTADL_0398	90	2.6
thiamine-phosphate pyrophosphorylase ThiE	1087	halTADL_0473	74	0.4
ribonucleoside-diphosphate reductase, alpha subunit (NrdE)	965	halTADL_0884	91	0.9
oxidoreductase FAD-binding domain	672	halTADL_1014	80	3.8
FAD-dependent pyridine nucleotide-disulfide oxidoreductase	650	halTADL_2528	90	4.2
rhodanese-like protein	258	halTADL_2750	92	25
adenine phosphoribosyltransferase (Apt)	985	halTADL_2952	95	0.9
dodecin	238	halTADL_3198	95	28
Oxidative stress				
ferritin Dps family protein	20	halTADL_1068	92	170
UspA domain-containing protein	851	halTADL_1904	76	1.6
glutaredoxin	1014	halTADL_2104	94	0.6
UspA domain-containing protein	910	halTADL_2276	75	1.2
thioredoxin	665	halTADL_2563	91	3.9

manganese/iron superoxide dismutase (Sod)	164	halTADL_2687	95	39
Protein chaperones				
group II chaperonin (thermosome)	135	halTADL_0092	91	45
Hsp20-type chaperone	335	halTADL_0114	93	17
peptidylprolyl isomerase FKBP-type	77	halTADL_0251	89	81
chaperone protein DnaK	212	halTADL_0595	94	32
heat shock protein Hsp20	217	halTADL_0724	95	31
prefoldin, beta subunit (PfdB)	177	halTADL_1114	95	37
group II chaperonin (thermosome)	19	halTADL_1928	95	200
prefoldin, alpha subunit (PfdA)	148	halTADL_2197	95	43
peptidylprolyl isomerase, cyclophilin type	152	halTADL_2273	91	42
peptidylprolyl isomerase, FKBP-type	827	halTADL_3026	90	1.8
group II chaperonin (thermosome)	6	halTADL_3279	95	387
Proteolysis				
membrane metalloprotease (peptidase M50)	640	halTADL_0323	79	4.4
proteasome alpha subunit (PsmA)	50	halTADL_2681	95	102
proteasome beta subunit (PsmB)	441	halTADL_2911	92	10
Ribosomes				
ribosomal protein L11	705	halTADL_0103	94	3.2
ribosomal protein L1	246	halTADL_0105	99	26
acidic ribosomal protein P0-like protein	478	halTADL_0106	94	8.7
ribosomal protein L7Ae	93	halTADL_0166	98	63
ribosomal protein S28e	922	halTADL_0167	95	1.2
ribosomal protein S7	304	halTADL_0623	87	19
ribosomal protein S6e	612	halTADL_2119	91	4.9
ribosomal LX protein	471	halTADL_2196	97	9.0
ribosomal protein S4	648	halTADL_2772	93	4.2
ribosomal protein L18e	713	halTADL_2775	93	3.0
ribosomal protein L13	607	halTADL_2776	96	5.0
ribosomal protein S3Ae	592	halTADL_3142	98	5.2
ribosomal protein S8e	762	halTADL_3327	89	2.5
ribosomal protein L30P	456	halTADL_3366	94	9.7
ribosomal protein L32e	268	halTADL_3370	91	24
ribosomal protein L6P	834	halTADL_3371	90	1.7

ribosomal protein S8	647	halTADL_3372	97	4.3
ribosomal protein L5	674	halTADL_3374	90	3.7
ribosomal protein S4e	338	halTADL_3375	90	16
ribosomal protein L29	194	halTADL_3380	92	34
ribosomal protein S19	643	halTADL_3383	97	4.3
ribosomal protein L23	463	halTADL_3385	88	9.3
ribosomal protein L4P	620	halTADL_3386	89	4.8
ribosomal protein L3	1001	halTADL_3387	93	0.8
Transcription				
TATA-box-binding protein (Tbp)	523	halTADL_0042	99	7.4
DNA-directed RNA polymerase subunit H (RpoH)	639	halTADL_0616	84	4.4
DNA-directed RNA polymerase subunit A (RpoA1)	589	halTADL_0619	96	5.3
DNA-directed RNA polymerase subunit A2 (RpoA2)	360	halTADL_0620	93	14
TATA-box-binding protein (Tbp)	966	halTADL_1732	96	0.9
DNA-directed RNA polymerase subunit D (RpoD)	499	halTADL_2774	91	7.9
DNA-directed RNA polymerase subunit N (RpoN)	842	halTADL_2778	97	1.6
Transcriptional regulators				
transcriptional regulator, AsnC family	921	halTADL_0058	93	1.2
phosphate uptake regulator, PhoU	699	halTADL_1186	92	3.3
transcriptional regulator, RosR (PadR family)	376	halTADL_1645	92	14
transcriptional regulator, XRE family	926	halTADL_2533	70	1.2
phosphate uptake regulator, PhoU	908	halTADL_3204	93	1.3
transcriptional regulator, AsnC family	172	halTADL_3422	96	38
Transduction				
response regulator receiver domain + HalX domain	229	halTADL_0055	96	29
response regulator receiver protein	566	halTADL_1808	94	6.0
KaiC domain	924	halTADL_1815	94	1.2
chemotaxis signal transduction protein CheW	925	halTADL_1838	91	1.2
Translation (other)				
translation elongation factor aEF-2 (FusA)	89	halTADL_0647	95	66
translation initiation factor 2, alpha subunit (a/eIF2-alpha) (Eif2a)	1090	halTADL_0923	91	0.4
translation initiation factor 2, beta subunit (a/eIF2-beta) (Eif2b)	545	halTADL_2337	95	6.7
methionyl-tRNA synthetase (MetG)	611	halTADL_3069	87	4.9
elongation factor 1-beta (aEF-1beta) (Ef1b)	454	halTADL_3453	94	9.7

Transport				
phosphate ABC transporter solute-binding protein (PstS)	85	halTADL_1182	51	71
ABC-type antimicrobial peptide transport system, permease component	693	halTADL_1613	88	3.4
iron ABC transporter solute-binding protein	856	halTADL_1788	83	1.6
carbohydrate ABC transporter solute-binding protein	247	halTADL_2761	84	26
carbohydrate ABC transporter ATPase	1099	halTADL_2764	88	0.4
branched-chain amino acid ABC transporter solute-binding protein	202	halTADL_2916	88	33
K+ uptake system, TrkA subunit	1048	halTADL_3061	89	0.6
nitrate/sulfonate/bicarbonate ABC transporter solute-binding protein	697	No match to tADL – unique to tADL-II	-	3.3

Table S10 Variant proteins detected for *Hht. litchfieldiae*, tADL, DL31 and *Hrr. lacusprofundi*. Protein numbers are given according to Table S3 (ranked by the sum of the normalized total spectrum count). Sequence identity refers to the amino acid sequence identity of the detected protein to its best match (column denoted "Locus tag") in a BLASTP search. Spectrum count shows the sum of the normalized total spectrum count across all 15 samples. SNP denotes variation caused by a single nucleotide polymorphism.

Protein annotation	Protein #	Locus tag	Sequence identity (%)	Spectrum count	Additional notes
***Hht. litchfieldiae* - variants**					
α-amylase (glycosyl hydrolase, family 13)	543	halTADL_0142	94	6.8	also detected for tADL and tADL-II
adhesion pilin (PilA)	750	halTADL_0751	65	2.6	
cell surface glycoprotein (Sec signal, PGF-CTERM, C-terminal transmembrane helix)	8	halTADL_1043	51	338	
cell surface glycoprotein (Sec signal, PGF-CTERM, C-terminal transmembrane helix)	18	halTADL_1043	44	205	
cell surface glycoprotein (Sec signal, PGF-CTERM, C-terminal transmembrane helix)	339	halTADL_1043	50	16	
cell surface glycoprotein (Sec signal, PGF-CTERM, C-terminal transmembrane helix)	356	halTADL_1043	42	15	
cell surface glycoprotein (Sec signal, PGF-CTERM, C-terminal transmembrane helix)	618	halTADL_1043	34	4.8	
hypothetical protein (TAT signal)	800	halTADL_1047	59	2.1	
SMC domain protein	709	halTADL_1458	62	3.1	
phosphate ABC transporter solute-binding protein (PstS)	61	halTADL_2155	88	92	
phosphate ABC transporter solute-binding protein (PstS)	75	halTADL_2155	93	82	
tADL - variants					
phosphate ABC transporter solute-binding protein (PstS)	13	halTADL_2155	99	235	SNP
methylated-DNA-[protein]-cysteine S-methyltransferase (Ogt)	273	halTADL_0579	99	24	SNP
winged helix-turn-helix DNA-binding domain	128	halTADL_0044	99	48	SNP
hypothetical protein	495	halTADL_2296	99	8.0	SNP
hypothetical protein (transmembrane helix near N-terminal)	852	halTADL_2505	97	1.6	
ribosomal protein L1	78	halTADL_0105	99	81	
methionyl-tRNA synthetase (MetG)	325	halTADL_3069	99	17	SNP

threonine synthase (ThrC)	655	halTADL_2266	97	4.0	
DL31 - variants					
RND superfamily / MMPL (mycobacterial membrane protein large) family protein	29	Halar_1791	99	134	
ribosomal protein L29	271	Halar_2474	99	24	
cell surface glycoprotein (Sec signal, PGF-CTERM, C-terminal transmembrane helix)	2	Halar_0829	47	629	
cell surface glycoprotein (Sec signal, PGF-CTERM, C-terminal transmembrane helix)	36	Halar_0829	45	118	
***Hrr. lacusprofundi* - variants**					
hypothetical cell surface protein (TAT signal)	27	Hlac_0476	49	138	
archaellin FlaA or FlaB	30	Hlac_2557	38	133	
cell surface glycoprotein (Sec signal, PGF-CTERM, C-terminal transmembrane helix)	223	Hlac_2976	38	30	

Table S11 Detected proteins with high sequence identity to *Hht. litchfieldiae,* DL31 or *Hrr. lacusprofundi* encoded on contigs with neighbouring genes that best matched to other haloarchaeal species. Protein numbers are given according to Table S3 (ranked by the sum of the normalized total spectrum count). Sequence identity refers to the amino acid sequence identity of the detected protein to its best match (column denoted "Locus tag") in a BLASTP search. Spectrum count shows the sum of the normalized total spectrum count across all 15 samples. Contig IDs are from the Deep Lake metagenome assemblies (Antarctic Lakes Metagenome: whole_lake.gbk at http://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=AntLakMetagenome).

Protein annotation	Protein #	Locus tag	Sequence identity (%)	Spectrum count	Contig ID
glycerol kinase (GlpK)	191	halTADL_0681	98	35	7180000420363
glycerol kinase (GlpK)	593	halTADL_0681	98	5.2	7180000399115
glycerol kinase (GlpK)	586	halTADL_0681	98	5.4	7180000397231
transcriptional regulator, AsnC family	47	halTADL_1491	88	104	7180000456564
hypothetical protein (Sec signal, Ig fold domain, C-terminal transmembrane helix)	163	halTADL_1042	73	39	7180000422580
transcriptional regulator, AsnC family	222	halTADL_1491	80	30	7180000396862
adhesion pilin (PilA)	358	halTADL_1885	33	15	7180000459513
hypothetical protein (transmembrane helix near N-terminal)	708	halTADL_1615	70	3.1	7180000398936
nucleic acid binding OB-fold tRNA/helicase-type (RPA32 homolog)	747	halTADL_3434	44	2.6	7180000455564
response regulator receiver protein	759	halTADL_2696	65	2.5	7180000268900
FeS assembly protein SufB	820	halTADL_0973	95	1.8	7180000432870
core histone	903	halTADL_1708	89	1.3	7180000446839
winged helix-turn-helix DNA-binding domain	909	halTADL_0044	62	1.2	7180000394553
SecD/SecF/SecDF export membrane protein	958	halTADL_0787	51	0.9	7180000457612
TATA-box-binding protein (Tbp)	980	halTADL_1450	93	0.9	7180000443701
Hsp20-type chaperone	795	Halar_3162	47	2.1	7180000457612
adhesion pilin (PilA)	839	Halar_2364	39	1.6	7180000242587
hypothetical protein (Sec signal; 2 x PKD/chitinase domains)	585	Hlac_2824	34	5.4	7180000456692
hypothetical protein: 3 x chitinase/PKD domains, C-terminal transmembrane helix	809	Hlac_2824	27	1.9	7180000434565
carbohydrate ABC transporter solute-binding protein	870	Hlac_2862	72	1.5	7180000267581
VCP-like protein (2 x CDC48 domains + 2 x AAA family ATPase domains)	1106	Hlac_2377	67	0.4	7180000414114

archaellin FlaA or FlaB	239	Hlac_2557	77	28	7180000295546