PNAS
www.pnas.org

"*Human subtlety will never devise an invention more beautiful, more simple or more direct than does Nature, because in her inventions, nothing is lacking and nothing is superfluous*."
Leonardo da Vinci (1452–1519)

SUPPLEMENTARY INFORMATION FOR

Symbiosis between nanohaloarchaeon and haloarchaeon is based on utilization of different polysaccharides

Violetta La Cono, Enzo Messina, Manfred Rohde, Erika Arcadi, Sergio Ciordia, Francesca Crisafi, Renata Denaro, Manuel Ferrer, Laura Giuliano, Peter N. Golyshin, Olga V. Golyshina, John E. Hallsworth, Gina La Spada, Maria C. Mena, Alexander Y. Merkel, Margarita A. Shevchenko, Francesco Smedile, Dimitry Y. Sorokin, Stepan V. Toshchakov, Michail M. Yakimov

Corresponding author. Email: mikhail.iakimov@cnr.it

This PDF file includes:

SI Extended Datasets S1, S2 and S3 (provided as separated files)

Supplementary Information: Text

Genome characteristics of nanohaloarchaeon LC1Nh and haloarchaeon LC1Hm

The genome of *Ca.* Nanohalobium constans LC1Nh consists of a single circular chromosome of 973,463 bp with molGC content 43.2%. The chromosome contains single copies of 5S-, 16S-, and 23S rRNA genes located in three different loci, as well as 39 tRNA genes, 23 of which have an intron. Of the 1,162 protein-coding genes annotated in strain LC1Nh, only 392 (i.e., 33.7% of the total) could be assigned to one of the NCBI COG (1) categories, and 732 (i.e., 63.3% of the total) to an arCOGs (2,3) (Tables S2 and S4). As of March 2020, there were 19 nanohaloarchaeal genomes of different degree of completeness deposited in NCBI and JGI databases but not included into COG and arCOG resources, and comparisons to these genomes were made as part of our analysis.

The genome of the host haloarchaeon *Halomicrobium* sp. LC1Hm (Table S3) consists of a circular chromosome of 3,105,114 bp with the GC content 65.7%, two divergent rRNA operons with two dissimilar *rrn*A and *rrn*B 16S rRNA genes (93.5% of gene identity), 48 tRNA genes and two CRISP repeat regions. Of the 3,318 protein-coding genes annotated in LC1Hm, 2,972 were assigned to arCOGs (91.1%). Additionally, *Halomicrobium* sp. LC1Hm had a circular megaplasmid of 223,917 bp with 64.1% GC content, encoding a single rRNA operon with *rrn*A 16S rRNA gene. This intraspecific polymorphism of 16S rRNA genes is typical of all known members of the genus *Halomicrobium*. The type species of the genus, *H. mukohataei*, has similar genome structure, similar arrangement of rRNA operons (4) and similar plasmid with 183 protein-coding genes. According to the presence of multiple genes encoding endochitinases of the GH18 family and the presence of β-*N*-acetylglucosaminidase of the GH20 family (5,6), the type species DSM 12286^T seems also to be a chitinotroph, although this was not proven by growth experiments prior to our study. We obtained this strain from DSMZ culture collection and after three passages of growth on the *Laguna Chitin* (LC) liquid medium, initially amended only with cellobiose (2 g l^{-1}), and then with cellobiose and amorphous chitin (2 g l^{-1} each of substrates), the disappearance of chitin was observed, indicating degradation of the polysaccharide.

Reconstruction of *Halomicrobium* sp. LC1Hm central metabolism

Halomicrobium sp. LC1Hm is an obligate aerobic heterotroph, capable of growth on chitin as

the sole source of carbon and energy. Cells from an axenic culture of the isolate were not able to use any of the other polysaccharides tested (cellulose, glycogen, starch or xylan). Phenotypic traits consistent with heterotrophy were confirmed using genome analysis, by finding the full sets of genes encoding for the glycolysis and gluconeogenesis, citrate cycle (TCA cycle), the pentose-phosphate pathway and pyruvate metabolism. Additionally, using the Kyoto Encyclopedia of Genes and Genomes (KEGG) as a reference database, fructose-, mannose-, sucrose-, amino sugar-, and nucleotide sugar-metabolism were reconstructed (SI Extended Dataset S1a-S1d). Although neither fructose nor mannose were tested as carbon sources for *Halomicrobium* sp. LC1Hm, this organism possesses various genes that are involved in their metabolism, and generate important substrates for N-glycan biosynthesis, amino- and nucleotide sugar metabolism and pyruvate metabolism. No fructose-specific transporter was detected, but an ABC transporter for glucose and mannose was identified (LC1Hm_1122-1124), indicating that *Halomicrobium* sp. LC1Hm can take up and assimilate these sugars from the environment.

As stated above, the experiment with axenic culture demonstrated that *Halomicrobium* sp. LC1Hm is not able to directly utilize complex polysaccharides such as cellulose, glycogen, starch and xylan, whereas it can utilize chitin. These data were further confirmed by genome annotation. In fact, a complex set of 28 CAZymes genes, including seven types of glycosyl hydrolase (GHs) of the class III endochitinases of GH18 family (EC3.2.1.14), and one GH20 family protein LC1Hm_0809, annotated as β-N-acetylglucosaminidase (EC3.2.1.14), which could hydrolyze chitodextrins and produce N-acetyl-β-glucosamine (GlcNAc), were identified (Table S8). Genes involved in hydrolyzing glycogen or starch were not detected. Besides the glucose-/mannose transporter mentioned above, one additional disaccharide-specific transporter (LC1Hm_0421-0424) was identified. Interestingly, while neither cellulase (EC 3.2.1.4) nor xylobiase (EC3.2.1.37) were found in the LC1Hm genome, cellobiase (EC 3.2.1.21, LC1Hm_1210) involved in the metabolism of cellobiose, was identified. Indeed, the axenic culture of *Halomicrobium* sp. LC1Hm grew on this disaccharide as the single source of carbon- and energy.

In our cultivation experiments, strain *Halomicrobium* sp. LC1Hm grew both in aerobic and microoxic conditions, but failed to prosper under anaerobic conditions. Although genes encoding different enzymes capable of incomplete nitrate reduction were identified; NarGHI nitrate reductase/nitrite oxidoreductase (EC 1.7.5.1) (LC1Hm_1872–1876), ferredoxin-nitrite

reductase (EC 1.7.7.1) (LC1Hm_2961), nitrite reductase (nitrite oxide-forming) (EC 1.7.2.1) (LC1Hm_0218) and nitric oxide reductase (EC 1.7.2.5) (LC1Hm_1303, neither respiration on nitrate nor single transformation of nitrate to nitrite did not occurred. The aerobic lifestyle of *Halomicrobium* sp. LC1Hm was confirmed genetically by the identification of all types of respiratory complexes: NADH-quinone oxidoreductase complex I (type Nuo, [EC 7.1.1.2]) (LC1Hm_1238–1248), complete succinate dehydrogenase / fumarate reductase (EC 1.3.5.1, EC 1.3.5.4) (LC1Hm_1406-1409), cytochrome c oxidase (EC 7.1.1.9) (LC1Hm_1283, LC1Hm_1293, LC1Hm_1658 and LC1Hm_2143), and cytochrome bd terminal ubiquinol oxidase (EC 7.1.1.7) (LC1Hm_1212-1213), as well as the V/A-type H^+/Na^+-transporting ATPase (EC 7.1.2.2) (LC1Hm_0489-0496).

Metabolism of both types of nucleotide resulted to be complete, demonstrating an apparent nucleotide prototrophy. Full sets of genes encoding enzymes involved in the biosynthesis and metabolism of all principal amino acids were also detected, also revealing an apparent amino-acid prototrophy (SI Extended Dataset S1a-S1d). A branched-chain amino-acid transport system (LC1Hm_1816–1820 and LC1Hm_1053) was detected, indicating on capability of *Halomicrobium* sp. LC1Hm to uptake leucine, isoleucine and valine from environment. Although we did not check the vitamin requirement for the growth of *Halomicrobium* sp. LC1Hm, the genes involved in thiamine-, riboflavin-, porphyrin-, nicotinate- and nicotinamide metabolism were identified together those involved in with pantothenate-, coenzyme-A-, folate-, ubiquinone- and terpenoid-quinone biosynthesis. Nevertheless, the LC media used for our cultivation-based studies was supplemented with vitamins (for all enrichment experiments, axenic culture of *Halomicrobium* sp. LC1Hm, and co-cultures of *Ca.* Nanohalobium constans LC1Nh and *Halomicrobium* sp. LC1Hm. Complex central metabolism of *Halomicrobium* sp. LC1Hm is supported by a full set of genes encoding for enzymes involved in genetic information-processing pathways, such as transcription, translation, folding, sorting, degradation, and replication and repair. All subunits of archaeal DNA polymerase B and D and archaeal RNA polymerase assure the complete/accurate replication and transcription of DNA, while a complete set of genes involved in the aminoacyl-tRNA biosynthesis, make this organism autonomous for translation and protein synthesis.

The protein export seems to be directed by the Sec-dependent pathway and by the twin-arginine translocation (Tat) system. Different genes involved in base-excision repair, nucleotide-excision repair and mismatch repair were also identified. These repair systems are

necessary to assure the correct functioning of replication processes needed to protect the cell from the variety of environmental stresses, such that caused by the UV-induced production of free oxygen radicals. We noticed a few genes coding for enzymes involved in the homologous recombination and non-homologous end-joining recombination along with the presence of at least two CRISPRs (clustered regularly interspaced short palindromic repeats) systems. The CRISPR1 system is positioned in the 1,018,721-1,022,107 region, and the CRISPR2 system is positioned from 2,022,402 to 2,023,289 with more than 51 and 13 spacer/repeaters identified, respectively. It is well known that the CRISPR systems protect the genome against plasmids, viruses and other potential mobile agents that could endanger genomic integrity. It is noteworthy, that a LC1Hm chromosome apparently harbours an intact phage of 11.1 Kb (from 16,781 to 26,879) (SI Extended Dataset S3d and S3e).

Finally, as we noticed by electron microscopy and by Nile Blue A staining, accompanied by fluorescence microscopy, during growth on chitin, each cell of *Halomicrobium* sp. LC1Hm produced numerous (intracellular) polyhydroxyalkanoate (PHA) granules. In agreement with these observations, we found genes LC1Hm_0239 and LC1Hm_0240, coding for two synthases, PhaC and PhaE, belonging to the Class III of PHA synthases (7), the enzymes catalysing the polymerization of PHA precursors (hydroxyacyl-CoAs like acetyl-CoA, propionyl-CoA, etc.). The PhaEC complex of *Halomicrobium* sp. LC1Hm is composed of two subunits: the catalytically active subunit PhaC (estimated pI = 3.77; molar mass 51,592 Da) and the structural subunit PhaE (estimated pI = 3.72; molar mass 19,911 Da), which is also indispensable for polymerization. In all known PHA-accumulating haloarchaea, PhaEC synthases of the class III polymerize short hydroxyacyl-CoAs, namely those no longer than 3-hydroxyvaleryl-CoA (see [8] for further information). No other biosynthesis enzymes, such as 3-ketothiolase and/or NADH/NADPH-dependent acetoacetyl–CoA reductase, typically needed for PHA biosynthesis in bacterial strains, were found in the *Halomicrobium* sp. LC1Hm genome. As it reviewed elsewhere, this simplified synthesis of PHA occurred via recruitment of only one type of biocatalyst is a particular feature of haloarchaea (8).

Limited anabolic capability of *Ca.* Nanohalobium constans LC1Nh.

Similar to most of the currently available DPANN genomes (9) and those of two recently cultivated Antarctic nanohaloarchaea (10), the 973,463 bp-long genome of *Ca*. Nanohalobium

constans LC1Nh is characterized by the absence of genes encoding the enzymes of canonical anabolic pathways necessary to synthesize most metabolic precursors and intermediates, including purines, pyrimidines, amino acids, cofactors and lipids. The LC1Nh genome is also missing pivotal enzymes needed for the pentose phosphate and ribulose monophosphate pathways, making it incapable of metabolizing pentose sugars.

In relation to genes need for metabolism of purine and pyrimidine, we identified only genes encoding kinases involved in the inter-conversion of nucleoside phosphates (LC1Nh_0358 and LC1Nh_0845) in the *Ca*. Nanohalobium constans LC1Nh genome. The enzymatic suite for *de-novo* biosynthesis of amino acids is also incomplete. There are genes present coding for only a fraction of the enzymes in this category, including asparagine synthase (LC1Nh_0884), aspartate aminotransferase (LC1Nh_0076), threonine dehydratase (LC1Nh_0072), chorismate mutase (LC1Nh_0074 and LC1Nh_0078), prephenate dehydratase (LC1Nh_0075) and prephenate dehydrogenase (LC1Nh_0077). These enzymes are mainly involved in the downstream stages of amino acids synthesis, or their salvage. Similarly, the ability for *de-novo* synthesis of cofactors is virtually absent, with only a few genes completing the synthesis or maturation of the most-common cofactors: nicotinamide-nucleotide adenylyltransferase (LC1Nh_0141), riboflavin synthase alpha chain (LC1Nh_0414), and lipoate-protein ligase (LC1Nh_1030). Enzymes involved in C_1 turnover are: dihydrofolate reductase (LC1Nh_0153) and 4α-hydroxytetrahydrobiopterin dehydratase (LC1Nh_0691). A caveat in this and other reconstructions is that one-third of the proteins encoded in the *Ca.* Nanohalobium constans LC1Nh genome cannot be assigned to any functional category and are annotated as hypothetical proteins, raising the question of whether this nanohaloarchaeon might encode some novel enzymes that drive canonical or entirely new metabolic pathways. However, the obligatory host-associated lifestyle of *Ca.* Nanohalobium constans LC1Nh, experimentally validated in this study, is consistent with the apparent paucity of important anabolic enzymes. We suspect that the same phenomenon occurs in other DPANN-host associations (11,12,13,14,15), nanohaloarchaea must acquire numerous metabolites from the host that are essential for their own metabolic functions. The same proposition of obligate host-dependence was made for two strains of *Ca*. Nanohaloarchaeum antarcticus that were co-cultivated with *Halorubrum lacusprofondi* (10).

Reconstruction of the central metabolism of *Ca.* Nanohalobium constans LC1Nh; expanded text related to Figure 5.

Protein translocation systems, membrane-associated proteases and transporters. Among a total of 1,205 proteins predicted from *Ca.* Nanohalobium constans LC1Nh genome, 36 proteins were annotated by BlastKOALA (16) as putative members of the 'Membrane Transport' category. Similar to the host (see above), two major pathways of protein secretion were found, apparently, complete, within the LC1Nh genome: the general secretion (Sec) and the twin arginine translocation (Tat) systems. The Sec secretory machinery includes the signal recognition particle complex (SRP-Sec), and five different genes belonging to this protein-trafficking system were recognized: SecE (LC1Nh_0103), SecY (LC1Nh_1093), SecF/D (LC1Nh_1168-9), the SRP receptor FtsY (LC1Nh_0658), and the targeting protein Ffh (LC1Nh_0652). The Tat pathway is another protein-transport system that exports folded proteins from cells of *Ca.* Nanohalobium constans LC1Nh. The TatA (LC1Nh_0432) and TatC (LC1Nh_0433) proteins were identified in the LC1Nh genome. Signal peptidases of SppA type (LC1Nh_0849) and of archaeal type I (LC1Nh_0018, 0300 and 0308) are the principal intra-membrane peptidases responsible for processing most of exported proteins in *Ca.* Nanohalobium constans LC1Nh cells. Two rhomboid-family proteases (LC1Nh_0186. 0519) and archaeosortase (LC1Nh_0663) are the intra-membrane serine proteases that cleave other proteins, including S-layer glycoprotein (LC1Nh_0029, 0824 and 1061) within their transmembrane domains. A total of 14 genes were annotated as components of ABC-type transporter systems. Among them are genes encoding: a substrate-binding protein (LC1Nh_0508), putatively involved in iron complex ABC transporter system; a putative peptide ABC transport system of SalY superfamily (LC1Nh_0028, 0030-31); and three uncharacterized ABC-2 type transporter complexes (LC1Nh_0314-16, LC1Nh_0707-10 and LC1Nh_0762-64). Sugars may be imported into the cytoplasm by one of these ABC-type transporters and/or by major facilitator superfamily (MFS) permease LC1Nh_0802. This putative transporter seems to be highly specific to nanohaloarchaea; we were unable to find any close homologs with annotated function. Unlike all currently recognized MFS permeases, which have 12 transmembrane helices (TMHs) organized in two domains within a single polypeptide chain, all putative permeases identified in the available nanohaloarchaeal genomes possess only 6-10 TMHs. The sodium-dependent phosphate co-transporter (LC1Nh_0626), zinc/iron permeases of the ZIP family (LC1Nh_0011, 1043) along with

potassium-dependent sodium-calcium exchanger (LC1Nh_0051), Na^+/K^+: proton antiporter of Kef type (LC1Nh_0771-2, 1088), NADH-dependent potassium transport system of Trk type (LC1Nh_0791-2), P-type heavy metal (cations)-transporting ATPase (LC1Nh_0696 and 1022), and K+-dependent mechano-sensitive channel (LC1Nh_1038) likely participate in inorganic ions transportation, osmotic homeostasis and resistance to heavy metals.

Cell-surface structures. The *Ca.* Nanohalobium constans LC1Nh genome has 20 genes likely encoding for the archaella assembly machinery and filament proteins (Table S7), and 11 of these genes (LC1Nh_0345-55) are organized in one operon hosting archaeal flagellar proteins FlaG, FlaI, FlaJ, FlaJ2 and seven hypothetical proteins and resembling the structure of euryarchaeal archaellum operons (17). Scanning electron microscopy of *Ca*. Nanohalobium constans revealed the presence of pilli-like structures similar in appearance and size to protein stalks of the archaella, which can unwind to thin filaments (Fig. S8 C,D,E). It still remains to be determined whether these flagellar structures are used for motility of detached *Ca*. Nanohalobium constans cells, or if they perform an alternative role in attachment to polysaccharide substrates and/or the host cell. Transmission and scanning electron microscopy revealed the spatial distribution of up to 17 nanohaloarchaeal cells on a single host cell, as well as what appears to be membrane stretching at the point of contact between the two organisms, suggesting strong intercellular interaction (Figs. S6 and S7). Given this, we looked for genomic signatures of the mechanisms by which these organism-organism interaction might occur. Besides archaella-related proteins, the *Ca*. Nanohalobium constans LC1Nh genome encodes 11 different glycosyl transferases belonging to families GH1, GH2 and GH8, the majority of which appear to be expressed according to proteomic data (SI Extended Dataset S2a and S2b). This indicates that nanohaloarchaeon LC1Nh must expend substantial amounts of ATP on the biosynthesis of precursors for glycosylation processes and synthesis of polysaccharides and glycoproteins as principal components of the extracellular matrix. Moreover, we detected at least eight strain-specific secreted proteins, containing polycystic kidney disease, (PKD)- (LC1Nh_0257, 0417, 0486, 0919) and concanavalin A-like/lectin- (LC1Nh_0399-401, 0423) LamG domains. Proteins that contain such domains serve a variety of purposes, often mediating interactions with carbohydrates or glycosylated proteins, and predicted to be likely involved in surface interactions in DPANN organisms (13,18,19).

Diversity-generating retroelements (DGRs). This family of genetic elements is known to modify DNA sequences and create massive sequence variation in targeted proteins typically involved in surface attachment, defence and regulation (20). The classical diversity-generating retroelements (DGR) mechanism of mutagenic homing deploys the error-prone reverse transcriptase (epRT) to modify the sequence of the copies of a target protein through diversifications of a RNA intermediate. Genes encoding the modified protein contain a variable repeat (VR) in close proximity to an invariant template repeat (TR); epRT-induced mutation of TR-RNA adenines and cDNA replacement of VR leads to an extraordinary degree of sequence diversification. DGRs occur widely in genomes of bacteria and their viruses, and seem to be prevalent in several DPANN lineages, as per the Genome Taxonomy Database (GTDB), such as *Nanoarchaeota*, including orders *Pacearchaeales* and *Woesearchaeales*, but not in *Nanohaloarchaeota* before our study (9,21). DGRs might play roles in cell–cell attachment, and providing DPANN organisms with a versatile tool for protein diversification that could enable adaptation to a (dynamic) host-dependent existence by conferring host specificity (21,22).

The DGRs locus was found in the *Ca.* Nanohalobium constans LC1Nh genome and its enzymatic part consists of the accessory variability determinant Avd (LC1Nh_0125) and the epRT (LC1Nh_0126) (Fig. S12). Between the Avd and epRT genes, the LC1Nh genome contains a 95-bp long TR, similar to the variable regions (VRA and VRB) of two proteins, LC1Nh_0008 and LC1Nh_0123 (80% and 84%, respectively). At the 3'-end of the VRA, VRB and TR regions, the LC1Nh DGRs system composes of three identical 19 bp-long sequences, coined as initiation of mutagenic homing sequences (IMH). Additionally to these core components, two hairpin/cruciform structures downstream of the VRA and VRB were evident in the LC1Nh genome. DGRs hairpin structures were seen to increase the efficiency of genetic information transfer from the TR to the variable regions through an RNA intermediate, a process termed as retro-homing (23). The DGR variable proteins LC1Nh_0008 and LC1Nh_0123 are very similar one to another and, as predicted by Phyre2 (24; http://www.sbg.bio.ic.ac.uk/phyre2), belong to the formylglycine-generating enzyme (FGE) subclass with a C-type lectin (CLec)-fold. This finding is consistent with the recent observation of remarkable conservation in archaea of the ligand-binding CLec-fold for accommodation of the massive sequence variation created by DGRs (22,25). The phylogenetic analysis of this

DGR locus revealed the presence of very similar DGRs-associated genes in the genome of *Haloplanus salinus* (Bioproject PRJNA481615). The *Haloplanus salinus* DGRs system possesses a much lower molGC% content (46.9%), compared with the genomic median molGC% of this heterotrophic haloarchaeon (66.2%) and is located within predicted genomic island, suggesting its acquisition by horizontal gene transfer from nanohaloarchaea.

ROS sensing and redox homeostasis.

As discussed in the main text, *Ca.* Nanohalobium constans lacks all respiratory complexes so must have a strictly anaerobic fermentative lifestyle. However, the mandatory dependence on aerobic host *Halomicrobium* sp. LC1Hm necessitates an extended tolerance of anaerobic LC1Nh to an oxygenated environment. Also as discussed in the main text, the host-ectosymbiont co-culture consumed twice as much oxygen as the *Halomicrobium* pure culture, producing microaerobic conditions at the early stationary phase of growth. Increased consumption of oxygen may be due to an increase in the metabolic needs of the haloarchaeal host, although the activity of non-respiratory oxygen-scavenging defence systems of nanohaloarchaeon could also contribute to the elimination of extra oxygen. Analysis of the LC1Nh genome revealed the presence of a thioredoxin system, consisting of two FAD-dependent thioredoxin reductases TrxR (LC1Nh_0509, 0593), one peroxiredoxin Prx (LC1Nh_0816) and three thioredoxins Trx (LC1Nh_0147, 0362, 0828). This sophisticated NAD(P)H-dependent redox system for thiol/disulphide cellular homeostasis in *Ca.* Nanohalobium constans cells may be a part of survival strategy under oxygen exposure, essential for adaptation to aerotolerance. Another finding is the presence in the LC1Nh genome of a putative NADH (peroxi)oxidase (LC1Nh_1142).

It is unclear whether this enzyme produces H_2O or H_2O_2, but its homolog has been shown to participate in oxygen scavenging and the regeneration of NAD in aero-tolerant anaerobic lactic acid bacteria, which lack the respiratory chains (26). We also found the superoxide dismutase SOD (LC1Nh_0512) and peptide-methionine sulfoxide reductase system composed of MsrA and MsrB, specific to each of the enantiomeric forms of the methionine sulfoxide (LC1Nh_0754 and 0961, respectively), potentially serving to counteract the effect of reactive oxygen species (ROS) produced during metabolism. In addition, we found four predicted extracellular dithiol-disulfide isomerases/oxidoreductases (LC1Nh_0053, 0067, 0469 and 0814) of DsrA/C family of thioredoxin proteins. In both prokaryotic and

eukaryotic organisms these proteins typically participate in oxidative protein folding via disulphide bond formation, breakage and isomerization (27). To perform the folding of extracellular proteins, the DsbA/C isomerases/oxidoreductases should be kept in different redox states by interaction with specific membrane-integrated DsbB and DsbG redox regulators. In the *Ca*. Nanohalobium constans LC1Nh genome we found the LC1Nh_0604 protein, which possessed five trans-membrane domains and was annotated as a disulfide bond formation protein of the DsbB family.

Energy production and catabolism.

The main genomic features of *Ca*. Nanohalobium constans LC1Nh are consistent with the predictions made on the basis of genome analyses of uncultured and cultured nanohaloarchaea, and other DPANN organisms (9,10,18,19). Besides significant genome reduction and presence of a full set of genes for chromosome maintenance, these include inability of synthesizing most necessary metabolic precursors, including amino acids and lipids, nucleotides and co-factors. The LC1Nh nanohaloarchaeon lacks xenorhodopsin genes, that are found in other nanohaloarchaeal genomes (28) and contains no evidence for carbon-fixation pathways, pointing at a strict heterotrophic lifestyle. Given that LC1Nh also lacks genes encoding known components of the tricarboxylic acid (TCA) cycle and any of the respiratory complexes (NADH dehydrogenase, functional cytochrome oxidases and terminal reductases), we infer a strictly anaerobic fermentation-based lifestyle. As the membrane-bound proton-translocating pyrophosphatases were also not detected, the maintenance of a chemiosmotic membrane potential (proton motive force) should rely on the A-type ATPase (LC1Nh_0829-37), the Kef-type potassium-hydrogen antiporter (LC1Nh_0771-2, 1088) and possibly other unidentified systems functioning as the outward proton-translocating membrane pumps.

Glycolysis. A complete set of genes for the archaeal type of dissimilative Embden-Meyerhof-Parnas (EMP) pathway of glucose degradation was identified in the *Ca*. Nanohalobium constans LC1Nh genome. At the same time, it lacks the Entner-Doudoroff pathway and both oxidative and non-oxidative variants of the pentose-phosphate pathway, found in some nanohaloarchaeotes (10,28,29,30). In the absence of membrane respiratory complexes, the EMP pathway is the only way of gaining energy by substrate-level phosphorylation. This

central pathway of energy production in nanohaloarchaea shows variations of the upper part also found in some methanogens, including halophilic members, but missing in haloarchaea (31). Employing ADP as the phosphoryl donor, the phosphorylation of both glucose and fructose 6-phosphate is likely catalysed by only one enzyme, bifunctional ADP-dependent phosphofructokinase/glucokinase LC1Nh_0114 of the ribokinase family (EC 2.7.1.146; EC 2.7.1.147). Fructose-1,6-biphosphate is further converted via fructose-biphosphate aldolase LC1Nh_0150 to dihydroxyacetone phosphate and glyceraldehyde-3-phosphate (GAP), which enter the lower portion of the EMP pathway and are further transformed by glyceraldehyde-3-phosphate dehydrogenase and phosphoglycerate kinase (LC1Nh_0135, 0188), ending with phosphoenolpyruvate (PEP). Of all archaea, only a few sugar-utilizing haloarchaea and glycogen-forming methanogens utilize this pair in the glycolytic direction for oxidation of glyceraldehyde-3-phosphate and ATP generation (32). The *Ca.* Nanohalobium constans LC1Nh genome hosts genes coding for three enzymes that can catalysing the final step of glycolysis, i.e. conversion of PEP to pyruvate: a pyruvate kinase (LC1Nh_0586) and two AMP/Pi-dependent phosphoenolpyruvate synthases (PEPS, LC1Nh_0232, 1145). Phosphorylations of glucose and fructose-6-phosphate by the phosphofructokinase/glucokinase require two molecules of ADP and produce two molecules of AMP. Assuming the use of the PGK/GAPDH enzyme pair and pyruvate kinase (LC1Nh_0586), four ATPs are generated from four ADPs. Adenylate kinase (LC1Nh_0358, 0845) regenerates these ADPs from two AMPs and two ATPs, resulting in the energy gain of two ATP molecules, which is standard for glycolysis in prokaryotes. However, the joint action of bifunctional ADP-dependent phosphofructokinase/glucokinase in the upper part and PEPS in the final step of this modified EMP pathway for sugar degradation could be energetically more favourable (33,34), since produced AMP can be directly re-consumed together with P_i, to form ATP, thus resulting favourable energy gain of four ATP molecules. As we mentioned above, the genome of *Ca.* Nanohalobium constans encodes a complete archaeal type ATPase complex (nine subunits, LC1Nh_0829-37). Thus, besides many important anabolic and homeostatic reactions, the energy that is produced can be used for maintenance of cytoplasmic pH within a biocompatible range and for providing a membrane potential by actively pumping protons out of the cytoplasm through the F_0 rotor of the A-type ATPase, thus resembling the metabolism of strictly fermentative organisms that lack electron transport chains and incapable of respiration.

Pyruvate metabolism. As many DPANN organisms, *Ca.* Nanohalobium constans LC1Nh uses the pyruvate dehydrogenase (PDH) complex LC1NH_0054-57 to decarboxylate pyruvate and to form acetyl-CoA. This complex is frequently found in aerobic bacteria and eukarya, but only a few sugar-utilizing hyperthermophilic and halophilic archaea possess it (35). ADP-forming acetyl-CoA synthetase (LC1Nh_0059) likely terminates the oxidative pathway of pyruvate metabolism with generation of ATP and formation of acetate. Oxidation of glucose to pyruvate involves the reduction of NAD^+ to NADH and thus, to avoid stopping glycolysis, the cells of *Ca.* Nanohalobium constans LC1Nh may have to re-oxidize the metabolically unused excess of this reduced electron/energy shuttle. There are multiple indications from the LC1Nh genome that pyruvate could be also used in the reductive pathway as an electron acceptor via NADH-dependent reduction by either lactate dehydrogenase (LC1Nh_0514), or NAD-dependent malic enzyme (LC1Nh_0063), or short-chain alcohol dehydrogenase (LC1Nh_0599, 1170).

Gluconeogenesis and glycogen metabolism. The *Ca.* Nanohalobium constans genome encodes a complete set of enzymes for the archaeal type of gluconeogenesis, including the diagnostic for this pathway bifunctional fructose-1.6-bisphosphate aldolase/phosphatase (LC1Nh_0149). The genome also contains genes coding for all key enzymes for glycogen synthesis: phosphomannomutase / phosphoglucomutase (LC1Nh_0113), glycogenin-like protein responsible for the initiation of the glycogen chain (LC1Nh_1199), UTP-glucose-1-phosphate uridylyltransferase (LC1Nh_1188), and glycogen synthase (LC1Nh_0117). This indicates that glucose formed *de novo* can be stored intracellularly in form of glycogen, besides being used for glycosylation of various membrane constituents and other metabolic needs. This type of carbon- and energy storage predicted to be a hallmark of DPANN organisms (9,18,19), but is absent in all known members of extremely halophilic archaea. Only a few methanogens, possessing the GAPDH/PGK enzyme-pair for oxidation of glyceraldehyde-3-phosphate, can store glucose in the same way (32). The capability for glycogen synthesis must be advantageous for *Ca.* Nanohalobium constans LC1Nh, because it provides at least temporary energetic independence to the cells that dissociate from their host. To decompose glycogen into glucose, LC1Nh encodes glycosyl hydrolases, such as glycogen debranching enzyme (GDE) / amylo-alpha-1,6-glucosidase (LC1Nh_0116), glucan-1,4-alpha-glucosidase / glucoamylase (LC1Nh_0129), and alpha-amylase (LC1Nh_0131).

Remarkably, all these glycosyl hydrolases were found in both the intracellular and extracellular proteomes (Table S7; SI Extended Dataset S2a and S2b), indicating that cells of *Ca.* Nanohalobium constans might also utilize glycogen from the environment. Joint action of these hydrolases outside the cells would ensure the cleavage of both alpha-1,4 and alpha-1,6 glycosidic linkages, present in exogenous glycogen, producing glucose extracellularly. As in case with GlcNAc, glucose can be transported inside the cells by dedicated sugar transporters (the LC1Nh genome contains three ABC-type transporters of unknown function LC1Nh_0314-6, LC1Nh_0707-10 and LC1Nh_0762-64).

Host-ectosymbiont interactions.

Our ability to grow nanohaloarchaea in the laboratory allowed us to elucidate key aspects of the trophic network between *Ca.* Nanohalobium constans LC1Nh and *Halomicrobium* sp. LC1Hm. As discussed in the main text, the phenotypic hallmark of the binary co-culture is its stable proliferation on insoluble chitin as a growth substrate. While the general analysis of the *Halomicrobium* sp. LC1Hm genome is presented above, here we analysed its chitinolytic genomic repertoire in more details. As is characteristic for *Halomicrobium mukohataei*-related strains, the LC1Hm haloarchaeon contains high number of CAZymes genes, including 26 various glycosyl hydrolases (GHs) (Table S8). Among them, seven GHs were unambiguously assigned to class III endochitinases of GH18 family (EC 3.2.1.14). All of them are predicted by SignalP 4.0 to have the N-terminal secretion signals, and all of them contain the ChtBD3 chitin-binding domains, suggesting that they may be able to attach to extracellular chitin particles and digest them. According to the CAZy classification, this type of endochitinases break down chitin microfibrils at internal sites forming low molecular weight chitodextrins/oligosaccharides $(GlcNAc)_{2\text{-}6}$. The *Halomicrobium* sp. LC1Hm genome also encodes a GH20 family protein LC1Hm_0809, annotated as β-*N*-acetylglucosaminidase (EC 3.2.1.14), which could hydrolyze chitodextrins and produce *N*-acetyl-β-glucosamine (GlcNAc). Similar to enochitinases, this enzyme may be acting outside the cells, as indicated by the presence of the signal peptide and two chitin-binding ChtBD3 domains at the N-terminus. The completely extracellular hydrolysis of chitin to GlcNAc by *Halomicrobium* sp. LC1Hm was confirmed by chromatographic analysis of its culture supernatant. Concentrations of GlcNAc were in range of 3.7-5.2 mmol in both chitin-grown axenic culture and in the *Ca.* Nanohalobium constans LC1Nh: *Halomicrobium* sp. LC1Hm co-culture. Such a high amount

of exogenous GlcNAc produced by *Halomicrobium* sp. LC1Hm is likely to be advantageous for its nanohaloarchaeal consort, since *Ca.* Nanohalobium constans has full set of enzymes which transform this monosaccharide into fructose-6-P that enters central carbohydrate metabolism. GlcNAc may be imported into the ectosymbiont cytoplasm by one of three ABC-type transporters found in its genome and/or by major facilitator superfamily (MFS) permease (LC1Nh_0802). Upon import into the cytoplasm, ATP-dependent kinase (LC1Nh_0180, 1157) can phosphorylate GlcNAc producing GlcNAc-6-P. During two consequent transformations catalyzed by glucosamine-phosphate N-acetyltransferase (LC1Nh_1149, 1150) and glucosamine--fructose-6-phosphate aminotransferase (LC1Nh_0637), GlcNAc-6-P is transformed into F-6-P, thus fueling both glycolysis and gluconeogenesis. The phosphorylation of GlcNAc apparently has no energetic costs, since consumed ATP can be regenerated from oxidation of acetyl-CoA by acetyl-CoA synthetase (LC1Nh_0059) leading to formation of acetate.

Another intriguing feature of *Ca.* Nanohalobium constans is the presence of (according to the genome) at least two extracellular serine proteases (LC1Nh_0159, 0909) and seven different cytoplasmic oligopeptidases (LC1Nh_0029, 0032-35, 0065, 0474) that might utilize exogenous peptides to generate amino acids used to supplement amino acid auxotrophy (Table S7). This proteolytic suite includes LC1Nh_0035 glutamyl aminopeptidase of M42 family of metallopeptidases. The same superfamily includes other hydrolases, such as cellulases and endo-1.4-beta-glucanases; this is perhaps the reason why this protein has been annotated in *Ca.* Haloredivivus, *Ca.* Nanopetramus, *Ca.* Nanosalina and *Ca.* Nanosalinarum spp. genomes as cellulase (28,29). Our cultivation experiments, however, showed that *Ca.* Nanohalobium constans was unable to grow on cellulose, arguing against the cellulolytic activity of this enzyme. However, the presence of the extracellular proteases might explain the ability of *Ca.* Nanohalobium constans to penetrate the cell wall of its haloarchaeal host observed by the thin section electron microscopy.

Lateral gene transfer between *Halomicrobium* and *Ca*. Nanohalobium

The absence of the majority of enzymes of canonical anabolic pathways suggests extensive trafficking of nutrients, metabolites and precursors of biopolymers from *Halomicrobium* host to the nanohaloarchaeon. At the same time, these two tightly associated species do not appear to be involved in any significant degree of gene exchange. Apart of the ancestrally collinear

loci coding for ribosomal proteins, RNA polymerase, ATPase and pyruvate dehydrogenase operons, comparative analysis with ProteinOrtho (36) of LC1Nh and LC1Hm deduced proteomes revealed the occurrence of only a few clusters of orthologous proteins (encoded by not more than two to three adjacent genes), shared between ectosymbiont and host (SI Extended Dataset S3a). It should be noted that all of these proteins differ in terms of gene context (< 50% of amino-acid identity) and genome alignments revealed the absence of significant regions of synteny between host and symbiont, confirming that no recent transfer of chromosomal fragments have between the two (Fig. 6).

It is worth mentioning that the genomes of *Halomicrobium* sp. LC1Hm and *Ca*. Nanohalobium constans LC1Nh differ significantly in their GC contents (65.66 and 43.24 mol GC%, respectively). Thus, the recent lateral gene transfer between these associated organisms could also be detectable as an anomaly in GC content of particular region(s) and further analyzed by parametric methods (37). Five non-ribosomal genes with GC content greater than 50% were found in the *Ca*. Nanohalobium constans LC1Nh genome. The horizontal gene transfer (HGT) detection tool of T-REX online portal (38,39) showed that two of them, organized in one operon encoding toxin of ParE/RelB superfamily, could be classified as HGT candidates (SI Extended Dataset S3b and S3c). Similar operons were found in haloarchaea *Salinirussus salinus* (WP_159903594-6) and *Haloplanus aerogenes* (WP_121921635-6), but not in the genome of *Halomicrobium* sp. LC1Hm. Notably, in the genome of *Haloplanus aerogenes*, this ParE/RelB operon was linked with tyrosine-type recombinase/integrase (WP_121921634) and thus, could be characterized as a part of its mobilome.

Additionally, both host and symbiont genomes were checked for the presence of genomic islands (GI) by means of Islandviewer 4 (40). Only one putative genomic island was detected in *Ca***.** Nanohalobium constans, consisting of a single gene (LC1Nh_0457) annotated as Nanohaloarchaeota-specific hypothetical protein with no putative conserved domains identified. None of 16 genomic islands, predicted in both the genome of *Halomicrobium* sp. LC1Hm, showed any indication of specific relationship with nanohaloarchaeon or its phylogenetic neighbors. The GI-associated genes in question likely came from halophilic archaea and either represented integrated mobile elements or prophages (Figure S13; SI Extended Dataset S3d-S3f). Subsequently, the genomes were the subjects of searches for viruses/prophages using two different versions of PHAge Search

Tool (41,42). Two genomic fragments were identified as prophages in the *Ca*. Nanohalobium genome at positions 131,538 - 144,312 bp and 953,766 - 962,535 bp (Figure S14). Based on the absence of key proteins of viral structure (capsid, head, plate, tail, coat, etc.), DNA regulation genes (encoding integrase, transposase and terminase), and functional genes (such as those coding for lysin and/or bacteriocin), both fragments were assigned as the incomplete prophages with a low GC content (42.09% and 44.56% respectively). Using the same approach, all currently available nanohaloarchaeal MAGs were screened for comparative analysis. Among 19 MAGs tested, no prophages were found in seven MAGs, while more than one incomplete prophage (IPs) were detected in '*Ca*. Nanohaloarchaeum antarcticus' Nha-CHI (3 IPs), '*Ca*. Nanohaloarchaeum antarcticus' Nha-R1 (2 IPs) and '*Ca*. Nanopetramus' sp. SG9 (2 IPs) (SI Extended Dataset S3d and S3e). Only one nanohaloarchaeon AB578-D14 was found to contain an intact halovirus VNH-1 (43). None of these structures was common between all checked nanohaloarchaea and *Ca*. Nanohalobium constans LC1Nh.

Supplementary Information: Materials and Methods

Description of sampling activities. Sediment cores (top 5 cm), mainly composed of salt crystals, and brine samples were collected in June 2016 from final crystallizer pond Vasca #27 of the *Saline della Laguna* solar saltern system (Fig. S1). Brine samples for determining major ion concentrations were collected in 1000-ml dark polyethylene (DPE) vials and stored at 25°C. Alternatively, 100 ml of the samples were acidified to pH 2.0 with 100 mM of HNO_3, diluted to salinity 35 g l^{-1} and stored in DPE vials at 25°C prior to chemical analyses, as described elsewhere (44). Determinations of oxygen concentration and redox potential were carried out according to previously established protocols (44).

Culture enrichment and co-culture isolation. The '*Laguna Chitin*' mineral medium (LC medium) was prepared on the basis of ONR7a medium (45), modified by the addition of (final concentration, g l^{-1}): 200 NaCl; 30 $MgCl_2$; 0.5 NH_4Cl and 5 HEPES (4-[2-hydroxyethyl]-1-piperazineethanesulfonic acid). The pH was adjusted to 7.0 by addition of 1 M KOH. Chitin from shrimp shells (Sigma-Aldrich, catalog number C9213) was added at the final concentration 5 g l^{-1} to serve as growth substrate. After sterilization, the medium was supplemented with 1 ml l^{-1} acidic trace metal solution (46), 1 ml l^{-1} vitamin mix (46) and 50 mg l^{-1} yeast extract. Five grams of the surface (0-2 cm) sediments were mixed with 500 ml of collected brine, and enrichment cultures were initiated by addition of 500 ml of sterile LC medium. These enrichments were incubated at 40°C in tightly closed 1200 ml, glass serum bottle without shaking in the dark. Determinations of oxygen concentration and redox potential were carried out following previously established protocols (44). Amorphous chitin was prepared by dissolving 10 g of crystalline chitin in 200 ml of concentrated HCl on ice overnight followed by dilution in a large volume (5000 ml) of ice-cold distilled water. The pH of solution was further neutralised by addition of 85 g of NaOH to precipitate the chitin, followed by low-speed centrifugation (6,000 rpm, 20 min, 4°C) using refrigerated floor centrifuge 360291 (Beckman Coulter). The final suspension was obtained by adding of 100 ml seawater to precipitate and sterilized by autoclaving at 120°C for 20 min in closed bottles. Differences in growth rates and oxygen consumption during cultivation on chitin of both the *Halomicrobium* sp. LC1Hm axenic culture and binary co-culture were calculated with one-way ANOVA, using

SigmaStat software. The relative importance of each treatment group was determined using the Holm-Sidak pairwise multiple comparisons procedure.

Attempts to obtain an axenic culture of *Ca*. Nanohalobium constans. Experimental design applying the dilution-to-extinction approach followed by filtering of ectosymbiont from the host, originally proposed for *Ca*. Nanoclepta minutus, was used to test the possibility of maintaining *Ca*. Nanohalobium as an axenic monoculture. The last positive dilution was passed through a 0.45 µm pore filter (Sartorius Stedim Biotech) three times, and the filtrate was checked by a *Halomicrobium*-specific PCR to verify the purity of nanohaloarchaeon cell-suspension (monitored also by use of nanohaloarchaeon- and archaeon-specific primers; Table S10). Thereafter, 1 ml filtrate was mixed with 20 ml fresh LC medium, each supplemented by addition of a different substrate at 2 g l^{-1} (final concentration): N-acetylglucosamine, dextrose, maltose, cellobiose, sucrose, starch, glycogen and chitin. To account for the theoretical possibility that *Ca*. Nanohalobium constans required any metabolite(s) from the *Halomicrobium* host, an aliquot (1 ml) of filtered nanohaloarchaeal cell-suspension was inoculated into 20 ml of spent medium from a 1 week-grown co-culture, obtained by sequential filtration (with passage through 0.45-µm filter, than 0.22-mm filter, and than [twice] through 0.1-µm pore filters; Sartorius Stedim Biotech). Additionally, 1 ml of the filtered nanohaloarchaeal cell-suspension was inoculated into 20 ml of fresh LC medium supplemented with cell lysates, obtained from either pure *Halomicrobium* sp. LC1Hm culture or the binary co-culture. To obtain these cell lysates, both the axenic *Halomicrobium* sp. LC1Hm culture and *Halomicrobium* : *Ca*. Nanohalobium co-culture were grown for two weeks at 40°C in 20 ml LC medium supplemented with chitin (2 g l^{-1}, final concentration). After this time, biomass (approx. 100 µg) was collected by centrifugation at 13,000 × g for 15 min at 4°C. The pellets were resuspended in MilliQ water and sonicated using a Vibracell Bioblock Scientific 75115 Sonicator (Sonics & Materials) with three bursts each of 30 sec at cycle duty of 50%. All these growth experiments were done in triplicate, statically cultivated for three weeks at 40°C, and then examined using CARD-FISH and analysed by PCR. To verify whether physical cell-to-cell contact with *Halomicrobium* sp. LC1Hm is a prerequisite for proliferation of *Ca*. Nanohalobium, we carried out the following experiment (in duplicate). First, 5 ml of chitin-grown *Halomicrobium* axenic culture was mixed with 230 ml fresh LC medium supplemented with chitin (2 g l^{-1}) and the bacteria-specific antibiotics, vancomycin

and streptomycin (100 mg l^{-1} of each, final concentration), followed by static incubation at 40°C in the upper chamber of Millipore® Stericup® filtration system (Merck) (capacity 250 ml, pore size 0.22 µm). After one week, axenic growth of *Halomicrobium* was obtained and the lower chamber was inoculated with 10 ml of filtered (3 x 0.45 µm) nanohaloarchaeal cell-suspension and 10 ml of fresh minimal LC medium (no chitin and any other carbon-substrate added). Throughout three weeks of incubation at 40°C, a vacuum pump was briefly applied each day to transfer 5-10 ml spent medium from the *Halomicrobium* culture to the lower chamber. Survival and eventual propagation of *Ca*. Nanohalobium cells in the lower chamber was determined by weekly examination using taxon-specific PCR. For these purposes, 10 ml of culture was collected from the lower chamber, passaged through a 0.1-µm filter to isolate DNA.

Host-switching experiments. A series of defined co-cultures were constructed by combining filtration-separated *Ca*. Nanohalobium cells with each of eight polysaccharidolytic and sugar-utilizing haloarchaeota. The following isolates, belonging to all three orders of the class *Halobacteria* were used as potential hosts: chitinolytic *Natrinema* sp. HArcht2 and chitinolytic *Salinarchaeum* sp. HArcht-Bsk1 (both of order *Natrialbales*); sugar-utilizing haloarchaeon *Haloferax volcanii* DSM3757 (order *Haloferacales*) and five members of order *Halobacteriales*, amylolytic *Halorhabdus thiamatea* SARL4B, cellulolytic *Halomicrobium zhouii* JCM 17095^T, cellulolytic *Halomicrobium zhouii* HArcel3, chitinolytic *Halomicrobium mukohataei* JCM9738^T and chitinolytic *Halomicrobium* sp. HArcht3-1. All cultures were grown in the LC liquid medium supplemented with corresponding polysaccharides (at 2 g l^{-1}) or dextrose in case of *Haloferax volcanii* (at 2 g l^{-1}). To improve the growth of *Halomicrobium zhouii* JCM 17095^T and *Halomicrobium mukohataei* JCM9738^T on corresponding polysaccharides, cellobiose (1 g l^{-1}) was also added. Cultures (80 ml) were produced in 120-ml serum bottles by incubating for four days at 40°C in the dark, without shaking. Cell suspensions of *Ca*. Nanohalobium (1 ml of early stationary phase co-culture, produced in LC broth) were then added to growing cultures after being passed three times through 0.45µm filters. Incubation (same conditions) was carried out for one month, and these cultures were then monitored for the presence of nanohaloarchaea by taxon-specific PCR. The LC1Hm + LC1Nh co-culture together with filtered nanohaloarchaeal cell-suspension were used as positive and negative control, respectively.

Analyses of metabolites. Oligosaccharides were analyzed by high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) on a ICS3000 system (Dionex, Thermo Fischer Scientific Inc., Waltham, MA) consisting of an single gradient pump, an electrochemical detector with a gold working electrode and Ag/AgCl as reference electrode, and an autosampler (model AS-HV). This mode of electrochemical detection id quantitatively accurate to a level of about 2 ng l^{-1} (47). All eluents were degassed by flushing with helium. An anion-exchange Carbo-Pack PA-200 column (4 × 250 mm, Dionex) connected to a CarboPac PA-200 guard column (4 × 50 mm) was used at 30°C. The initial mobile phase was 4 mM NaOH at 0.3 ml min^{-1} for 30 min. Then the column was washed for 20 min at 0.5 ml min^{-1} with a solution containing 100 mM sodium acetate and 100 mM NaOH, and equilibrated with 4 mM NaOH. Acetate concentrations were measured by gas chromatography (Chromoteq-Crystall 5000.2; column Sovpol-5, 1 m, photoionization detector PID in the temperature range 180°C to 230°C) after cell removal by filtration through 0.1-µm filter and acidification of the supernatant to pH 4.0 with 2.0 M HCl. Additionally, the short-chain organic acids acetate, lactate and malate were determined in the supernatants obtained by centrifuging enrichment cultures using corresponding colorimetric assay kits (Sigma-Aldrich catalog numbers MAK086, MAK064 and MAK067-1KT), according to the manufacturer's protocols.

16S rRNA diversity analysis. Small aliquot (2 ml) of each chitinotrophic enrichment was centrifuged at 10,000 × g at 4°C for 10 min. Total DNA was extracted using a GNOME DNA kit (MP Biomedicals, USA), according to manufacturer's instructions. The quantity and purity of DNA was estimated by NanoDrop® ND-1000 Spectrophotometer (Wilmington, DE, USA), and the 16S rRNA gene was amplified using taxon-specific set of primers for Nanohaloarchaea to distinguish them from haloarchaea (Table S10). The reaction was carried out in LifeEco PCR Thermal Cycler (BIOER Technology, China) as follows: initial denaturation at 94°C for 5 min; 35 cycles of 1 min at 94°C, 1 min at 60°C, and 2 min at 72°C; final extension step of 10 min at 72°C. PCR products were visualized on a 1% (w/v) agarose gel. Amplicons for cloning were cut out from the gel and purified with the Wizard SV Gel and PCR Clean–up System kit (Promega, Madison, WI, USA). The haloarchaea present in the last positive enrichment (Fig. S1) were identified following the cloning and sequencing of the 16S rRNA gene using the conventional archaeal universal primers A20F-1492R (48,49). Library

construction was performed as described elsewhere (50) and 48 positive clones from each library were Sanger-sequenced by Eurofins Genomics (Ebersberg, Germany).

The V3-V4 hypervariable regions of SSU rRNA gene were PCR amplified and the amplicons were sequenced on the Illumina MiSeq platform by FISABIO, Valencia, Spain (http://fisabio.san.gva.es/en/servicios). Library preparations were performed as described previously (51). The 16S rRNA gene V3-V4 variable region was amplified using PCR primers 341F/785R S-D-Bact-0341-b-S-17, 5'-CCTACGGGNGGCWGCAG-3' (52), and S-D-Bact-0785-a-A-21, 5'-GACTACHVGGGTATCTAATCC-3' (52,53). The raw data generated this way were further filtered (direct removal of barcodes and primer sequences, sequences shorter than 150 bp, as well as sequences with ambiguous base calls and with homopolymer runs longer than 6 bp) by the same sequencing service. Pre-processed sequences were analysed with the NGS analysis pipeline of the SILVA rRNA gene database project (SILVAngs 1.3) (54). Each sequence was aligned using the SILVA Incremental Aligner (SINA v1.2.10) for ARB SVN (55) against the SILVA SSU rRNA SEED and quality-controlled database (54). After quality control steps, sequences were de-replicated on a per-sample basis. Identical reads were collapsed and the unique reads were clustered (OTUs 98% similarity) using cd-hit-est (version 3.1.2; http://www. bioinformatics.org/cd-hit) (56). The reference read of each OTU was classified by a local nucleotide BLAST search against the non-redundant version of the SILVA SSU Ref dataset (release 132, Dec 13, 2017; http://www.arb-silva.de) using Blastn (version 2.2.30+; http://blast.ncbi.nlm.nih.gov/Blast.cgi) with default settings (57). Reads assigned to Nanohaloarchaea were collected. Only OTUs made up by more than 50 reads were included in the phylogeny analysis and were filtered using the Python script filter_fasta.py (http://qiime.org/scripts/filter_fasta.html). For the initial phylogenetic tree, filtered reads and close sequence relatives were aligned using the SILVA alignment tool and manually inserted in ARB. The sequences were then re-aligned using the latest SILVA database for ARB (release 132 SSURef NR99). The initial 16S rDNA tree (Fig. S5) was constructed using phylogenetic inference under maximum parsimony criteria within the ARB software environment. One thousand bootstrap re-samplings were performed to estimate the robustness of the tree partitions.

DAPI and CARD-FISH analyses. Co-culture samples were fixed with formaldehyde (FA, 2% v/v final concentration) for 1 h at 25°C and cells were collected by slow filtration onto

polycarbonate membranes (25-mm diameter, 0.22-μm pore size, GE Healthcare Bio-Sciences, USA). Filters for counts of reporter deposition coupled with fluorescence *in-situ* hybridization (CARD-FISH) were embedded in 0.1% (w/v) low temperature-melting agarose (Sigma-Aldrich), dried at 37°C for 10 min, and dehydrated with 95% (v/v) ethanol. After eliminating the ethanol, cells were permeabilized with lysozyme (10 mg ml^{-1}, 1 h) and achromopeptidase (5 mg ml^{-1}, 30 min) at 37°C. Intracellular peroxidase was inhibited by treatment with 10 mM HCl at room temperature for 20 min. Following acid treatment, filters were washed with 0.22-μm-filtered MilliQ water, dipped in 95% (v/v) ethanol and air-dried. Filters were cut in sections prior to hybridization with the oligonucleotide probes. Using a previously published protocol for fluorescence *in-situ* hybridization (FISH) analysis applied for Nanohaloarchaea, we did not observe any significant hybridization signal. The Nanohaloarchaea-specific horseradish peroxidase probe Narc_1214 for CARD-FISH analysis (was synthesized by Biomers (Ulm, Germany). Detailed information about the probes is reported in Table S9. *Halomicrobium*-specific probe Hmb_172 was used for quantification of the host cells, since Nanohaloarchaea possess eight mismatches in the probe-specific site (5'-A**C**A**AT**GCA**C**T**A**ACG**TGC**G-3'). The horseradish peroxidase probes (2 μl each) were added from working solutions prepared at concentrations of 50 ng DNA $μl^{-1}$. Hybridization conditions were optimized at 46°C for 3h, as previously described (58). Amplification was performed at 46°C for 40 minutes using tyramide Alexa488 and Alexa594 for Narc_1214 and Hmb_172 respectively. After amplification, filters were washed in 1× PBS, rinsed in Milli-Q water, dehydrated in 96% ethanol and air-dried. Visualization of all cells was done with DAPI mix (4',6'–diamidino-2-phenylindole, DNA staining dye; final concentration 2 μg ml^{-1}) in a 4:1 ratio of Citifluor (Citifluor Ltd, Leicester, UK) and Vectashield (Linaris GmbH, Wertheim-Bettingen, Germany).

Staining of cells with Nile Blue A. As it described elsewhere (59), 0.22 μm-filtered aqueous solution of Nile Alue A (1%, w/v, Sigma-Aldrich catalog number V000845) was used to cover the dried cell suspension of archaeal cells, obtained by the heat fixation (55°C, 20 min) on a microscopic slide. The slide was carefully wrapped within moistened tissue paper, then inserted into a horizontal 50-ml Falcon tube, closed and stained by incubating the tube in a 55°C water bath for 15 min. After staining, the slide was washed with tap water to remove excess of stain solution, followed by washing with an 8% (v/v) acetic acid solution for 1 min,

washed again with tap water, and allowed to air-dry. After addition of Vectashield H-1000 antifading mounting medium, the slide was covered with a glass cover slip. The samples were then examined with the AXIOPLAN 2 Imaging microscope (Zeiss, Germany) and fluorescence was observed by illuminating with blue wavelengths.

Determination of cell numbers by quantitative PCR. The quantitative PCR (qPCR) method was used to determine relative cell densities of *Halomicrobium* sp. LC1Hm and *Ca*. Nanohalobium constans LC1Nh in co-culture grown in hypersaline LC mineral medium supplemented with vitamins, yeast extract, antibiotics and amorphous chitin, at pH 7.0 and 40°C. DNA for the qPCR was extracted from 2.0 ml of LC1Hm+LC1Nh co-cultures collected at fixed times, corresponding to lag, exponential and early stationary phases of growth using GNOME DNA kit (MP Biomedicals, USA). Extracted DNA was dissolved in 50 µl of TE buffer (10 mM Tris-HCl, 1 mM EDTA [pH 7.5]) and quantified using NanoDrop ND-1000 spectrophotometer (Euroclone, Milan, Italy). The quality of the extracted DNA was checked by electrophoresis in a 1.0% (w/v) agarose gel. The qPCR was performed with SYBR Green on an ABI Prism 7300 Real-time PCR System (Applied Biosystems, Foster City, CA, USA). All amplifications were checked for specificity with dsDNA melt curves and samples exhibiting multiple products were not considered in the analysis. Primers based on the sequences of 16S rRNA genes of LC1Hm and LC1Nh were designed using Primer Express software, version 2.0 (Applied Biosystems, Foster City, CA, USA) and listed in Table S10. All primers were checked for specificity using BLAST searches for short, nearly exact matches at NCBI. All of the sequences in GenBank database, except the desired targets, had ≥ 2 mismatches with the primers. Additionally, the primers were screened for specificity *in silico* using ProbeCheck (60). To obtain DNA standards for exact quantification, single clones containing either *Ca*. Nanohalobium constans LC1Nh or *Halomicrobium* sp. LC1Hm 16S rRNA genes in pGEM-T Easy Vector (Promega, Madison, WI, USA) were grown overnight at 37°C and plasmid was subsequently purified using the NucleoBond Xtra Midi kit (Macherey-Nagel). Plasmid DNA was solubilized in 150 µl of Tris-EDTA buffer and quantified using NanoDrop ND-1000 spectrophotometer. With the molecular weight of the plasmid and insert known, 10-fold serial dilutions of pGEM_Nh and pGEM_Hm were prepared, ranging from 3.76×10^3 to 3.76×10^9 copies μl^{-1} and from 1.07×10^3 to 1.07×10^9, respectively. Serial dilutions were prepared at the same time for each target and used for the qPCR in triplicate to create the

standard curve for sample quantification. Primer concentrations were chosen to minimize the length of quantification cycle, or C_q, of the standard, while also to minimizing primer-dimers and non-target amplification, as assessed though post-amplification dsDNA melt curves. Each 25 µl reaction contained 50 ng of DNA, 12.5 µl of SYBR Green Master Mix (ThermoFisher) and 200 nM of each primer. The qPCR protocol included the following steps: an initial denaturation step at 95°C for 10 min, followed by 45 cycles of denaturation at 95°C for 15 s and annealing/elongation at 60°C for 60 s. A dissociation step was added to check for primer-dimer formation. From the slope of each curve, PCR amplification efficiency (E) was calculated according to the following equation (61): $E = [10^{-1/slope}] - 1$. Obtained slope values fell within the optimal range corresponding to efficiencies of 99.6% and 96.2%, respectively.

Genome sequencing and assembly. Whole-genome shotgun sequencing of the *Ca*. Nanohalobium constans LC1Nh: *Halomicrobium* sp. LC1Hm co-culture was done by FISABIO (Valencia, Spain) using the Illumina® MiSeq System platform (San Diego, CA, United States) with 2 x 300 bp short insert paired-end libraries (MiSeq Reagent Kit v3). FISABIO also performed the quality assessment and the sequence joining (forward R1 and reverse R2). Quality assessment was performed with PRINSEQ-lite program (62) using the following parameters: Min_length:50bp trim_qual_right:30bp trim_qual_typ:mean, trim_qual_window:20bp. Forward and reverse reads were joined by the FLASH program (63) applying default parameters. The resulting reads (8,479,286 total paired in two runs, after performing quality check) were assembled by Unicycler program (64) with default (normal mode) parameters and using the joined sequences provided as long reads (-l command). Velvet 1.2.1058 plugin inside Geneious package, and the Geneious 7.1 built-in assembly software (Biomatters Ltd., New Zealand) were used to verify consistency and refine contigs generated from Unicycler, yielding about 147 x of Ca. Nanohalobium and 83 x of *Halomicrobium* genome coverage respectively. For closing the *Halomicrobium* sp. LC1Hm genome, a pure isolate was further sequenced using the MinION™ Oxford Nanopore Technologies platform (Oxford, UK), yielding the necessary long reads (11,197 total, with 10,371 bp of mean and 103,876 bp of the longest) to perform a hybrid assembly by Unicycler.

Gene prediction and annotation. Prediction of *Ca*. Nanohalobium constans LC1Nh protein-coding genes was performed by Glimmer 3.02 (65), and verified by the EMBOSS 6.5.7 (66) inside Geneious 7.1. The FgenesB server (67) (www.softberry.com) was used for operon

prediction. Predicted protein sequences were manually annotated by integrating the results of Blastx/Blastp searches against NCBI NR database with information got from other databases, including KEGG, Clusters of Orthologous Groups (COGs), and arCOGs (2,68,69) (http://www.genome.jp/tools/blast/), as well as PATRIC/RAST server (70), and BlastKOALA (16) (https://www.kegg.jp/blastkoala/). For prediction of tRNA and rRNA genes, tRNAscan-SE 2.0 online (71) and RNAmmer 1.2 online (72) tools were used. Carbohydrate-active enzymes, glycoside hydrolases in particular, were detected by using NCBI Blastp against the Carbohydrate Active Enzymes database (73) (CAZy, http://www.cazy.org/). Transmembrane prediction regions were performed by the TMHMM program (74) (http://www.cbs.dtu.dk/services/TMHMM/). N-glycosylation sites were predicted using the NetNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc/). Genomic comparisons were visualized using Circos software (75). Percentages of amino-acid identity levels used as the input for Circos visualization were obtained by the PATRIC/RAST server. Prediction of genomic islands was performed using IslandViewer4 (40) applying two different algorithms, IslandPath-DIMOB and SIGI-HMM, based on analysis of mobile elements and codon usage/dinucleotide distribution biases. All found genomic islands (GIs) were further verified manually by checking the taxonomic affiliation of best blast hits of predicted horizontally transferred proteins.

Protein-based phylogeny. The list of 122 archaeal core genes was taken from Genome Taxonomy DataBase (GTDB) (https://gtdb.ecogenomic.org/). These marker genes were identified in selected genomes, aligned and concatenated using GTDBtk v0.3.2 (76). Alignment consisting of 32,675 aa positions was automatically trimmed using trimAl 1.2rev59 (77). Phylogenetic tree was built using PhyML 3.0 program and the Bayesian-like transformation of approximate likelihood-ratio test for branch support (78). The substitution model for phylogenetic inference was automatically selected by SMS algorithm (79). Nanohalorachaeal genomes with a completeness level of < 65% were omitted from the analysis. The following non-nanohaloarchaeal genomes were included: Aenigmarchaeota (GCA_000806115.1, GCA_002688965.1, GCA_002784265.1, GCA_002789635.1, GCA_002791855.1 and UBA10154); Micrarchaeota (GCA_000830275.1, GCA_000402355.1, GCA_002778455.1, GCA_002778455.1, GCA_001871475.1 and SRX764827); Nanoarchaeota (GCA_000008085.1 and GCA_001552015.1); Parvarchaeota

(GCA_002498685.1, GCA_002502045.1, GCA_002502525.1, GCA_002503215.1, GCA_002503305.1, GCA_002503915.1 and GCA_002503995.1) and Woesearchaeota (GCA_000830295.1, GCA_002688315.1, GCA_002688925.1, GCA_002762785.1, GCA_002762985.1, GCA_002779235.1, GCA_000806095.1, SRX764834, SRX764704, UBA10171, UBA12494). An alternative phylogeny by using concatenated ribosomal proteins, and inserting 16 of the 18 nanohaloarchaeal genomes currently deposited in NCBI database, was also performed (Fig. S6). For each genome, the amino acids sequences of 11 ribosomal proteins were aligned by Clustal W 2.1 program with BLOSUM substitution matrix (80). The aligned proteins were further concatenated in the following order: 30S ribosomal protein S2 (arCOG04245); 30S ribosomal protein S5 (arCOG04087); 30S ribosomal protein S7 (arCOG04254); 30S ribosomal protein S8 (arCOG04091); 30S ribosomal protein S9 (arCOG04243); 30S ribosomal protein S11 (arCOG04240); 50S ribosomal protein L1 (arCOG04289); 50S ribosomal protein L5 (arCOG04092); 50S ribosomal protein L6 (arCOG04090); 50S ribosomal protein L13 (arCOG04242); 50S ribosomal protein L18 (arCOG04088). The tree was inferred by PhyML 3.0 plugin inside Geneious 7.1 with Blosum62 substitution model and 1,000 bootstrap replicates (81). In this analysis, the cluster of haloarchaeal sequences was used as the out-group.

Field emission scanning electron microscopy (FESEM). The cells grown in the LC broth were fixed with 2% (v/v, final concentration) freshly prepared paraformaldehyde. Fixative was removed by washing twice with LC mineral medium before final fixation with aqueous osmium tetroxide (four parts LC mineral medium and one part 5% [w/v] aqueous osmium tetroxide) for 30 min at 25°C. The fixed material was washed with LC mineral medium and placed onto poly-L-lysine-coated cover slips for 10 min, followed by treatment with 1% (v/v) glutaraldehyde to cross-link microbes with poly-L-lysine coating. This step prevents washing away during the dehydration and critical-point drying of the attached microorganisms. Dehydrating was achieved using series of acetone water mixtures and pure acetone (10, 30, 50, 70, 90, and 100% [v/v] acetone) on ice for 10 min for each step. Once in the 100% acetone, samples were allowed to reach room temperature (evaporation of acetone has cooled the sample), replenishing with fresh 100% acetone. Samples were then subjected to critical-point drying with liquid CO_2 (CPD 030, Bal-Tec, Liechtenstein). Dried samples were covered with a gold-palladium film by sputter coating (SCD 500 Bal-Tec, Liechtenstein) before

examination in a field-emission scanning electron microscope Zeiss Merlin (Carl Zeiss, Oberkochen) using the Everhart Thornley SE-detector and the in-lens secondary electron detector in a 50 : 50 ratio with an acceleration voltage of 5 kV. Contrast and brightness were adjusted with Adobe Photoshop CS5.

Ultrathin sections and transmission electron microscopy (TEM). Glutaraldehyde-fixed samples (see FESEM Section) were washed twice with growth medium and fixed with 1% (w/v) aqueous osmium (final concentration) for 1 hour at room temperature. Samples were then embedded into hot aqueous agar solution 2% (w/v), solidified agar was cut into small cubes and dehydrated with a graded series of ethanol-water mixtures and pure ethanol (10, 30, 50, 70, 90, and 100% [v/v] ethanol) for 30 minutes at each step. The 100% ethanol step was repeated twice before the samples were infiltrated with one part 100% ethanol and one part LRWhite resin for 8 h followed by one part 100% ethanol and two parts LRWhite resin for 24 h hours, and subsequently infiltrated with pure LRWhite resin with two changes over 2 days. The next day 1 µl starter was added to 10 ml LRWhite resin, stirred and resin was put into 0.5 ml gelatin capsules. The samples were placed into the tip of the capsules, followed by polymerization for 2 days at 50°C. Ultrathin sections were cut with a diamond knife. Sections were counter-stained with 4% (w/v) aqueous uranyl acetate for 3 min. Samples were examined in a TEM910 transmission electron microscope (Carl Zeiss, Oberkochen) at an acceleration voltage of 80 kV. Images were taken at calibrated magnifications, recorded digitally with a Slow-Scan CCD-Camera (ProScan, 1024x1024, Scheuring, Germany) with ITEM-Software (Olympus Soft Imaging Solutions, Münster, Germany), and the contrast and brightness were adjusted with Adobe Photoshop CS5.

Negative staining of whole cells for TEM. Thin carbon support films were prepared by sublimation of a carbon thread onto a freshly cleaved mica surface. Osmium-fixed cells were adsorbed onto the carbon film and negatively stained with 0.5% (w/v) uranyl acetate solution, pH 5.0, according to the method of (82). Samples were examined in a TEM 910 transmission electron microscope (see above).

Protein digestion. After the 10 ml of chitin-grown *Ca*. Nanohalobium : *Halomicrobium* co-culture was centrifuged at 13,000 × g for 15 min at 4°C of, the cell pellet was treated with a lysis buffer containing 7 M urea (USB Corporation, Cleveland, OH), 2 M thiourea (Sigma-

Aldrich), 5% (w/v) CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate, Sigma-Aldrich), 5 mM TCEP (Sigma-Aldrich) and a protease inhibitor cocktail (Sigma-Aldrich), and incubated for 15 min on ice. Homogenization of the pellet was achieved by ultrasonication for 5 min on ultrasonic bath Branson 2510 (Marshall Scientific, New Hampshire, USA). The homogenate was centrifuged at 20,000 × g for 10 min at 4°C, and the supernatant containing the solubilized proteins was used for further analysis. The cell-free supernatant obtained after first step of centrifugation was concentrated using a Nanosep 10 kDa cut-off filter (Pall Corporation) to obtain the extracellular protein fraction. Then, 20 µg of proteins in each sample were precipitated by methanol/chloroform method and re-suspended in 30 µl of multichaotropic sample solution UTT buffer (7 M urea, 2 M thiourea, 100 mM triethylammonium bicarbonate buffer [TEAB, Sigma-Aldrich]) and 10 µl of 20% (w/v) sodium dodecyl sulfate was added. Each re-suspended sample was chemically reduced using 4 µl of 50 mM Tris(2-carboxyethyl)phosphine (TCEP, Sigma-Adrich), pH 8.0, at 37°C for 60 min, followed by addition of 2 µl of 200 mM cysteine-blocking reagent methyl methanethiosulfonate (SCIEX, Foster City, CA, USA) for 10 min at room temperature. Each sample was diluted to the final volume of 400 µl with S-Trap buffer (90% methanol [v/v], 100 mM TEAB) and digested into the S-Trap Micro Spin Column (ProtiFi, NY, USA), according to the manufacturer's instructions. Digestion was initiated by adding 1 µg Pierce Trypsin Protease, of MS-grade (Thermo-Fisher Scientific Inc.) to each sample in a ratio 1:20 (w/w), and then incubated at 37°C overnight on a shaker. Sample digestion was evaporated to dryness in a vacuum concentrator.

Liquid chromatography mass spectrometry (LC-MS) analysis. Digested sample was desalted using Stage-Tips with Empore 3M C18 disks (Sigma-Aldrich). A 1 µg-aliquot of resulting peptides obtained was subjected to the 1D-nano LC ESI-MS/MS analysis (liquid chromatography electrospray ionization tandem mass spectrometric) using a nano-liquid chromatography system Eksigent Technologies NanoLC Ultra 1D plus (SCIEX, Foster City, CA, USA) coupled to high-speed Triple TOF 5600+ System (SCIEX) with a NanoSpray III source. The analytical column used was a silica-based reversed phase ACQUITY UPLC⟍ M-Class Peptide BEH C18 Column (Waters Corporation, Milford, MA, USA). The trap column was a C18 Acclaim PepMap^{TM} 100 (Thermo-Fisher Scientific Inc.), 100 µm × 2 cm and with a 5-µm particle diameter, 100 Å pore size, switched on-line with the analytical column. The

loading pump delivered a solution of 0.1% formic acid in water at 2 µl/min. The nano-pump provided a flow-rate of 250 nL/min and was operated under gradient elution conditions. Peptides were separated using a 250 min gradient ranging from 2 % to 90 % mobile phase B (mobile phase A: 2% (v/v) acetonitrile [Scharlab S.L.], 0.1% (v/v) formic acid [Sigma-Aldrich]; mobile phase B: 100 % acetonitrile, 0.1% (v/v) formic acid). Injection volume was 5 µl. Data was acquired using an ion spray voltage floating 2300 V, curtain gas 35, interface heater temperature 150, ion source gas 1 25 and declustering potential 150 V. For IDA parameters, 0.25 s MS survey scan in the mass range of 350–1250 Da were followed by 35 MS/MS scans of 100 ms in the mass range of 100–1800 Da. Switching criteria were set to ions greater than mass to charge ratio (m/z) 350 and smaller than m/z 1250 with charge state of 2–5 and an abundance threshold > 90 counts (cps). Already-encountered target ions were excluded for 15 s.

Proteomics data analysis and sequence search. Mass spectrometry data were processed using PeakView v2.2 Software (SCIEX) and exported as .mgf files. The resulting mass spectra were searched against the predicted tryptic peptide sequences encoded by *Ca*. Nanohalobium constans LC1Nh and separately the predicted tryptic peptide sequences encoded by *Halomicrobium* sp. LC1Hm (9,394 sequences), together with commonly occurring contaminants, using the open-source software X!TandemPipeline version 3.4.3 (83). Search parameters were set as follows: enzyme, trypsin; allowed missed cleavages, 2; methylthiolation (C) as fixed modification and acetyl (Protein N-term), pyrrolidone from E, pyrrolidone from Q and oxidation (M) as variable modifications. Peptide mass tolerance was set to ± 25 ppm for precursors and 0.05 Da for fragment masses. The confidence interval for protein identification was set to ≥ 95% ($p<0.05$) and only peptides with an individual ion score above the 1% False Discovery Rates (FDR) at spectra level were considered. Only proteins identified with at least two unique peptides were considered as correctly identified. The protein abundance index (PAI) and exponentially modified protein abundance index (emPAI) were calculated as described elsewhere (84). The PAI = N_{obsd}/N_{obsbl}, where N_{obsd} and N_{obsbl} are the number of observed peptides per protein and the number of observable peptides per protein, respectively (85).

Supplementary Tables

Table S1. Major ion concentrations of the *Saline delle Lagune* brine sample from the crystallizer pond Vasca #27 (Fig. S1) compared with average Mediterranean seawater and seawater-derived brine (10-fold evaporated seawater). All concentrations are in g l^{-1}. Reported values for the Vasca #27 brine are mean ± 5% (n = 6).

Major ions	Mediterranean seawater[a]	Vasca #27 brine	Seawater-derived brine[a]	Ratio of Vasca #27 brine : seawater
Na^+	12.4	86.7	124.7	7.0
Cl^-	21.2	138.0	218.2	6.5
Mg^{2+}	1.48	9.8	12.6	6.6
K^+	0.47	3.5	4.0	7.5
Ca^{2+}	0.48	0.2	0.23	0.4
SO_4^{2-}	2.95	20.4	20.9	6.9
Br^-	0.069	0.50	0.63	7.2
Li^+	1.8×10^{-4}	1.24×10^{-3}	0.002	6.9

[a] These data taken from (50).

Table S2. General characteristics of the *Ca.* Nanohalobium constans LC1Nh genome.

Feature	Value
Chromosome size	973,463 bp
GC content	43.2%
Protein-coding regions (%)	883,599 bp (90.8%)
Total genes	1,204
tRNA genes	39 (23 introns)
rRNA genes (5S-16S-23S)	3 (in 3 different operons)
Protein-coding genes	1,162
Proteins assigned to COGs (%)	392 (33.7%)
Proteins assigned to arCOGs (%)	735 (66.3%)
Average gene length	760.4 bp
Max gene length	4,500 bp
ATG initiation codon proteins	1,014
GTG initiation codon proteins	111
TTG initiation codon proteins	37

940 **Table S3.** General characteristics of the *Halomicrobium* sp. LC1Hm genome.

Quality	Value
Chromosome size	3,105,114 bp
GC content	65.7%
Protein-coding regions (%)	2,736,699 bp (88.1%)
Total genes	3,318
tRNA genes	48 (3 introns)
rRNA genes (5S-16S-23S)	6
CRISPR regions	2
Protein-coding genes	3,264
Proteins assigned to COGs (%)	2,256 (69.1%)
Proteins assigned to arCOGs (%)	2,972 (91.1%)
Average gene length	833.1 bp
Max gene length	6,555 bp
ATG initiation codon proteins	2,530
GTG initiation codon proteins	665
TTG initiation codon proteins	69
Plasmid size	223,917 bp
GC content	64.1%
Protein-coding regions (%)	196,892 bp (87.9%)
Total genes	187
tRNA genes	-
rRNA genes (5S-16S-23S)	3
CRISPR regions	1
Protein-coding genes	183
Proteins assigned to COGs (%)	130 (71.0%)
Proteins assigned to arCOGs (%)	164 (89.6%)
Average gene length	1,029.7 bp
Max gene length	4,470 bp
ATG initiation codon proteins	138
GTG initiation codon proteins	43
TTG initiation codon proteins	2

945

Table S4 Genes of *Ca.* Nanohalobium constans LC1Nh assigned to various functional classes using the NCBI COG database.

Code	Value	Percentage	COG category	Function
J	92	7.92%	Translation. ribosomal structure and biogenesis	Informational
K	23	1.98%	Transcription	
L	43	3.70%	Replication; recombination and repair	
O	25	2.15%	Posttranslational modification; protein turnover; chaperones	Cellular
D	5	0.43%	Cell cycle control; cell division; chromosome partitioning	
M	25	2.15%	Cell wall/membrane/envelope biogenesis	
T	4	0.34%	Signal transduction mechanisms	
U	5	0.43%	Intracellular trafficking; secretion and vesicular transport	
V	17	1.46%	Defence mechanisms	
C	18	1.55%	Energy production and conversion	Metabolic
E	16	1.38%	Amino acid transport and metabolism	
F	15	1.29%	Nucleotide transport and metabolism	
G	16	1.38%	Carbohydrate transport and metabolism	
H	7	0.60%	Coenzyme transport and metabolism	
I	3	0.26%	Lipid transport and metabolism	
P	13	1.12%	Inorganic ion transport and metabolism	
R	32	2.76%	General function prediction	Uncharacterized
S	33	2.84%	Function unknown (uncharacterized)	
-	**770**	**66.26%**	**Not in COGs**	-

950

Table S5. *Candidatus* Nanohaloarchaeota whole genome sequencing (WGS) projects currently registered in the NCBI (JGI) Genome Database. Organisms deposited in Genome Taxonomy Database are marked with stars. Genomes that are completed (circular closed) genomes as at February 2020 are in bold.

Organism name	Accession	Scaffolds	Contigs	Length (bp)	% GC	Isolation Source	Note
***Ca.* Nanohalobium constans LC1Nh**	**CP040089.1**	**1**	**1**	**973,463**	**43.2**	**Italy: Mozia solar salterns. Trapani**	**Cultured**
Ca.* Nanopetramus sp. SG9	**CP012986.1**	**1**	**1**	**1,118,574**	**46.4**	**Chile: Salar Grande Halite rock**	**Metagenomic data**
***Ca.* Nanohaloarchaeum antarcticus Nha-CHI**	**Ga0310355 (JGI)**	**1**	**1**	**1,093,273**	**40.3**	**Antarctica: Hypersaline Lake Club**	**Cultured**
Ca. Nanohaloarchaeum antarcticus Nha-R1	Ga0101775 (JGI)	3	3	1,094,064	40.3	Antarctica: Hypersaline Lake Rauer 1	Enrichment
Ca. Haloredivivus sp. G17	AGNT00000000.1	448	448	1,198,604	42.0	Spain: Santa Pola salterns. Alicante	Single-cell isolation
Nanohaloarchaea archaeon AB578-D14	AYGT00000000.1	962	962	1,028,544	52.5	Spain: Santa Pola salterns. Alicante	Single-cell isolation
Ca. Nanosalinarum sp. J07AB56	AEIX00000000.1	3	259	1,215,802	55.4	Australia: Hypersaline Lake Tyrrell	Metagenomic data
Ca. Nanosalina sp. J07AB43*	AEIY00000000.1	7	210	1,227,157	43.6	Australia: Hypersaline Lake Tyrrell	Metagenomic data
Nanohaloarchaea archaeon B1-Br10_U2g1*	LKMN00000000.1	26	26	709,831	42.6	Russia: Kulunda steppe Lake Bitter-1 brine	Metagenomic data
Nanohaloarchaea archaeon B1-Br10_U2g19*	LKMO00000000.1	35	35	662,884	41.0	Russia: Kulunda steppe Lake Bitter-1 brine	Metagenomic data
Nanohaloarchaea archaeon B1-Br10_U2g21*	LKMP00000000.1	24	24	815,638	39.4	Russia: Kulunda steppe Lake Bitter-1 brine	Metagenomic data
Nanohaloarchaea archaeon B1-Br10_U2g29*	LKMQ00000000.1	35	35	526,182	40.1	Russia: Kulunda steppe Lake Bitter-1 brine	Metagenomic data
Nanohaloarchaea archaeon PL-Br10_U2g16*	LKMV00000000.1	54	54	652,532	42.2	Russia: Kulunda steppe Lake Bitter-1 brine	Metagenomic data
Nanohaloarchaea archaeon PL-Br10_U2g19*	LKMW00000000.1	55	55	758,571	42.4	Russia: Kulunda steppe Lake Bitter-1 brine	Metagenomic data
Nanohaloarchaea archaeon PL-Br10_U2g27*	LKMX00000000.1	52	52	581,882	43.2	Russia: Kulunda steppe Lake Bitter-1 brine	Metagenomic data
Nanohaloarchaea archaeon QH_8_44_6	PXPB00000000.1	329	345	565,289	44.1	Chile: Atacama Desert salt crust	Metagenomic data
Nanohaloarchaea archaeon SW_10_44_10	PXPC00000000.1	132	144	397,623	43.3	Chile: Atacama Desert salt crust	Metagenomic data
Nanohaloarchaea archaeon SW_4_43_9	PXPD00000000.1	51	71	661,323	43.4	Chile: Atacama Desert salt crust	Metagenomic data
Nanohaloarchaea archaeon SW_7_43_1	PXPE00000000.1	1	1	956,550	42.6	Chile: Atacama Desert salt crust	Metagenomic data
Nanohaloarchaea archaeon SW_7_46_7	PXPF00000000.1	191	224	572,429	45.7	Chile: Atacama Desert salt crust	Metagenomic data

Table S6. Median isoelectric point (pI) and amino-acids composition (%) of (annotated and predicted) proteins in *Ca*. Nanohalobium constans LC1Nh, the haloarchaeal host, *Halomicrobium* sp. LC1Hm and selected DPANN organisms with WGS projects registered into NCBI Genome Database. Organisms validated in GTDB Taxonomy are marked with stars. DPANN organisms, not belonging to Nanohaloarchaeota, are shown in blue.

Species	pI	A Ala	C Cys	D Asp	E Glu	F Phe	G Gly	H His	I Ile	K Lys	L Leu	M Met	N Asn	P Pro	Q Gln	R Arg	S Ser	T Thr	V Val	W Trp	Y Tyr
***Halomicrobium* sp. LC1Hm**	**4.47**	**11.0**	**0.7**	**8.7**	**8.3**	**3.2**	**8.4**	**2.0**	**3.9**	**1.7**	**8.9**	**1.7**	**2.2**	**4.6**	**2.8**	**6.7**	**5.6**	**6.7**	**9.1**	**1.2**	**2.7**
***Ca.* Nanohalobium constans LC1Nh**	**4.65**	**6.0**	**0.6**	**7.2**	**11.4**	**4.0**	**7.0**	**1.7**	**6.3**	**6.3**	**8.2**	**2.3**	**4.7**	**3.3**	**3.6**	**4.4**	**6.9**	**5.3**	**6.7**	**0.9**	**3.1**
Ca. Nanopetramus sp. SG9*	4.69	6.1	0.6	7.2	11.5	4.0	7.4	1.6	5.9	6.2	8.2	2.3	4.4	3.3	3.4	4.8	6.8	5.3	6.8	0.9	3.3
Ca. Haloredivivus sp. G17	4.84	5.8	0.7	6.6	10.8	4.1	6.7	1.6	6.5	6.3	8.7	2.6	4.4	3.3	3.4	5.3	7.3	4.9	6.4	0.9	3.2
Ca. Nanosalinarum sp. J07AB56	4.67	7.6	0.8	7.4	9.7	3.5	7.9	1.8	4.3	3.7	8.8	2.1	3.3	4.0	3.6	6.7	7.6	5.6	8.0	0.9	2.7
Ca. Nanosalina sp. J07AB43*	4.74	5.9	0.7	7.3	10.1	3.8	6.9	1.7	6.2	5.8	8.2	2.4	4.6	3.4	3.8	5.0	7.6	5.3	6.8	0.9	3.3
Ca. Nanohal. antarcticus Nha-CHI	4.84	5.6	0.7	7.1	9.9	3.9	6.9	1.7	6.8	6.3	8.5	2.5	5.0	3.4	3.8	4.6	7.1	5.4	6.6	0.9	3.3
Ca. Nanohal. antarcticus Nha-R1	4.75	5.7	0.6	7.3	10.1	4.0	7.1	1.7	6.9	6.4	8.4	2.2	5.0	3.3	3.7	4.5	7.1	5.2	6.6	0.9	3.3
Archaeon B1-Br10_U2g1*	4.74	5.8	0.6	7.0	12.0	3.8	6.8	1.7	6.4	6.4	8.6	2.3	4.1	3.3	3.2	4.9	6.5	5.2	7.2	0.8	3.3
Archaeon B1-Br10_U2g19*	4.76	5.7	0.7	7.1	11.9	3.9	6.7	1.7	6.3	6.5	8.6	2.2	4.4	3.3	3.3	4.6	6.6	5.4	7.1	0.8	3.4
Archaeon B1-Br10_U2g21*	4.70	6.1	0.7	7.2	11.7	3.9	6.9	1.7	6.6	6.0	8.5	2.3	4.6	3.3	3.3	4.5	6.7	5.0	6.7	0.9	3.4
Archaeon B1-Br10_U2g29*	4.75	6.0	0.6	7.1	11.4	4.2	6.9	1.6	6.7	6.2	8.5	2.5	4.5	3.3	3.3	4.6	6.8	4.9	6.9	0.9	3.2
Archaeon PL-Br10_U2g16*	4.65	5.5	0.6	7.7	12.1	3.9	7.1	1.6	6.5	6.5	8.3	2.4	4.2	3.2	3.3	4.7	6.8	4.7	6.9	0.8	3.2
Archaeon PL-Br10_U2g19*	4.79	5.6	0.7	7.0	12.0	3.9	6.7	1.7	6.4	6.5	8.5	2.3	4.2	3.5	3.2	4.9	6.5	5.3	7.0	0.8	3.4
Archaeon PL-Br10_U2g27*	4.90	6.4	0.8	6.7	11.3	4.0	6.9	1.7	6.4	6.4	8.9	2.4	4.1	3.4	3.3	4.8	6.7	4.9	6.9	0.9	3.2
Archaeon QH_8_44_6	4.90	5.5	0.9	6.1	10.5	3.8	6.7	1.8	6.0	6.7	8.2	2.3	4.3	3.4	3.7	6.4	7.1	5.3	6.1	1.0	3.1
Archaeon SW_10_44_10	4.88	5.6	0.8	6.3	10.8	3.7	6.8	1.7	6.4	7.0	8.1	2.5	4.4	3.3	3.5	5.8	6.9	5.1	6.4	0.9	3.1
Archaeon SW_4_43_9	4.81	5.7	0.8	6.6	11.2	3.8	7.0	1.6	6.5	6.8	8.2	2.3	4.6	3.3	3.4	5.3	6.9	5.2	6.5	0.9	3.2
Archaeon SW_7_43_1	4.79	5.7	0.7	6.9	11.3	3.9	7.0	1.6	6.6	6.7	8.2	2.3	4.9	3.2	3.3	4.8	6.9	5.2	6.6	0.9	3.3
Archaeon SW_7_46_7	4.81	5.8	0.8	6.3	10.5	3.7	7.1	1.8	5.7	6.3	8.1	2.3	4.2	3.5	3.9	6.3	7.1	5.4	6.3	0.9	3.0
Nanoarchaeon B72_G9	8.27	6.1	1.6	5.9	6.6	4.0	6.4	2.0	8.5	8.6	8.5	2.6	4.8	3.5	2.7	4.5	6.3	5.8	6.6	0.9	3.8
Nanoarchaeon B49_G9	8.84	6.0	1.6	5.2	7.0	4.7	5.9	1.7	9.7	10.2	8.4	2.6	5.7	3.2	2.6	3.8	6.2	5.0	5.5	0.8	3.9
Nanoarchaeum equitans Kin4-M*	8.81	5.2	0.8	5.0	7.9	4.4	5.3	1.3	10.5	10.8	10.4	1.7	5.3	4.0	2.2	3.9	4.6	4.1	5.9	1.0	5.6
Aenigmarchaeota ex4484_224*	9.30	4.5	0.9	3.9	8.8	5.7	5.5	1.2	10.7	12.2	10.1	1.8	4.6	3.4	2.1	4.5	5.8	3.5	6.0	1.1	3.8

Table S7. Proteins of the central metabolic and homeostatic reconstruction map depicted in Figure 5 and Figure 7 (see also main text).

Category	Locus tag	Gene product	Name in Figure 5	Proteome biomass	Proteome secretome
Proteases and peptitades	LC1Nh_0018,0300	signal peptidase I		YES	NO
	LC1Nh_0308	signal peptidase I		NO	NO
	LC1Nh_0032,0033,0035,0065,0159,0474	peptidase		YES	YES
	LC1Nh_0909	serine protease Do		YES	YES
	LC1Nh_0076	aspartate aminotransferase		YES	NO
Poly- and oligo- saccharides	LC1Nh_0116	glycogen debranching enzyme (alpha-1,6-glucosidase)	GDE	YES	YES
	LC1Nh_0129	glycosyl hydrolase family 15	GH15	YES	YES
	LC1Nh_0131	glycosyl hydrolase family 57	GH57	YES	YES
	LC1Nh_0795	glycosyl hydrolase family 1	GH1	YES	YES
	LC1Nh_0796	glycosyltransferase family 2 protein		YES	YES
N-acetyl glucosamine metabolism	LC1Nh_0180,1157	gluco(hexo)kinase	HK	YES	YES
	LC1Nh_0637	glucosamine--fructose-6-phosphate aminotransferase	GNFAT	YES	YES
	LC1Nh_1149	bifunctional UDP-N-acetylglucosamine pyroph.	GNPAT	NO	YES
	LC1Nh_1150	bifunctional UDP-N-acetylglucosamine pyroph.	GNPAT	YES	YES
	LC1Nh_0802	major facilitator superfamily MFS_1	MFS	YES	NO
Glycolysis	LC1Nh_0114	ADP-dependent phosphofructokinase/glucokinase	PF/GK	NO	YES
	LC1Nh_0115	triosephosphate isomerase (TIM)	TIM	YES	YES
	LC1Nh_0118	glucose-6-phosphate isomerase	PGI	YES	YES
	LC1Nh_0135	glyceraldehyde-3-phosphate dehydrogenase (NAD(P))	GAPD	YES	YES
	LC1Nh_0150	fructose-1,6-bisphosphatase I	FBPA	YES	YES
	LC1Nh_0188	phosphoglycerate kinase	PGK	YES	YES
	LC1Nh_0232	phosphoenolpyruvate synthase / pyruvate, water dikinase	PEPS	YES	YES
	LC1Nh_0282	enolase, C-terminal TIM barrel domain	ENO	YES	NO
	LC1Nh_0358,0845	adenylate kinase	AK	YES	NO
	LC1Nh_0586	pyruvate kinase	PK	YES	YES
	LC1Nh_0954	2,3-bisphosphoglycerate-independent phosphoglycer.	PGM	YES	NO
	LC1Nh_1033	glucose-1-phosphatase		YES	NO
Pyruvate Metabolism and Fermentation	LC1Nh_0054,0055,0056,0057	pyruvate dehydrogenase (dihydrolipoamide dehydr.)	PDH	YES	YES
	LC1Nh_0059	acetate-CoA ligase (ADP-forming) subunit alpha	ACL	YES	NO
	LC1Nh_0063	malate dehydrogenase (oxaloacetate-decarboxylating)	ME	YES	YES
	LC1Nh_0514	D-lactate dehydrogenase	LDH	YES	NO
	LC1Nh_0599,1170	short-chain alcohol dehydrogenase	ADH	YES	YES
Gluconeogenesis	LC1Nh_1033	glucose-1-phosphatase	G1P	YES	NO
	LC1Nh_0149	fructose-bisphosphate aldolase, class I	FBP	YES	YES
Glycogen synthesis	LC1Nh_0113	phosphomannomutase / phosphoglucomutase	PGM	YES	YES
	LC1Nh_0117	glycogen synthase	GS	YES	YES
	LC1Nh_1188	nucleotidyl transferase	GPUT	YES	YES
	LC1Nh_1199	glycogenin-like protein	GLP	YES	NO
Secretion	LC1Nh_1093,1168	preprotein translocase subunit SecY	Sec	NO	YES
	LC1Nh_1169	preprotein translocase subunit SecD	system	YES	YES
	LC1Nh_0432	sec-independent protein translocase protein TatA	Tat	YES	NO
	LC1Nh_0433	sec-independent protein translocase protein TatC	system	NO	NO

Oxidative stress	LC1Nh_0147	thioredoxin family protein	Trx	NO	NO
	LC1Nh_0362,0593	thioredoxin	TrxR	YES	NO
	LC1Nh_0509	thioredoxin reductase (NADPH)	TrxR	NO	NO
	LC1Nh_0512	superoxide dismutase, Fe-Mn family	SOD	YES	YES
	LC1Nh_0754	peptide-methionine (S)-S-oxide reductase	MSRA	NO	NO
	LC1Nh_0816	OsmC family peroxiredoxin	Prx	NO	NO
	LC1Nh_0828	glutaredoxin		NO	NO
	LC1Nh_0961	peptide-methionine (R)-S-oxide reductase	MSRB	YES	NO
	LC1Nh_1142	pyridine nucleotide-disulphide oxidoreductase	NOX	YES	NO
Cell surface & Flagella	LC1Nh_0029,0824,1085	S-layer domain protein		YES	YES
	LC1Nh_0305,0500,0738,0929	PilT protein domain protein		YES	NO
	LC1Nh_0350,1130	archaeal flagellar protein FlaI		YES	YES
	LC1Nh_0351,0352,1131	archaeal flagellar protein FlaJ		YES	YES
	LC1Nh_0941	type IV secretory pathway ATPase VirB11/Archaellum		YES	YES
	LC1Nh_1129	hypothetical protein		YES	YES
	LC1Nh_1132	archaeal flagellar protein FlaJ		NO	YES
	LC1Nh_0399,0400,0401,0423	concanavalin A-like lectin/glucanases superfamily	LamG	YES	YES
	LC1Nh_0791,0792	Trk-type K+ transport system, membrane component	TrkAH	YES	YES
	LC1Nh_1088	Kef-type K+ transport system, membrane component	KefBC	YES	YES
	LC1Nh_0626	Na+/phosphate symporter	Symportr	YES	NO
	LC1Nh_0008	formylglycine-generating enzyme	CLec-fold protein	YES	YES
DGR	LC1Nh_0123	hypothetical protein		YES	NO
	LC1Nh_0124	hypothetical protein		NO	NO
	LC1Nh_0125	diversity-generating retroelement protein bAvd	Avd	NO	NO
	LC1Nh_0126	RNA-dependent DNA polymerase (Reverse transcr.)	RT	YES	YES
	LC1Nh_0127	transcriptional regulator like protein		NO	NO
Energy	LC1Nh_0829-0837, LC1Nh_0833-0837	V/A-type H+/Na+-transporting ATPase subunits	A-type ATPase	YES	YES
Protein translocation and Posttranslational modification	LC1Nh_0284,0475,0554,0585	proteasome	Proteaso me	YES	YES
	LC1Nh_0849	protease IV		YES	NO
	LC1Nh_0186,0519	rhomboid family intramembrane serine protease		YES	YES
PKD-containing proteins	LC1Nh_0257,0486,0919	PKD domain containing protein	PKD	YES	NO
Protein-disulfide isomerase / oxidoreductase Dsb system	LC1Nh_0053,0067,0814	DSBA oxidoreductase	Dsb system	YES	YES
	LC1Nh_0605	disulfide bond formation protein DsbB		NO	NO

Table S8. Proteins identified by CAZy and NCBI databases as glycoside hydrolases (GH) in the *Ca.* Nanohalobium constans LC1Nh and *Halomicrobium* sp. LC1Hm genomes. Endochitinases and exochitodextrinase of *Halomicrobium* sp. LC1Hm are highlighted in yellow.

locus_tag	NCBI Accession	GH family	Activities in family (most common)	Blastp bitscore	Blastp E-value
LC1Nh_0116	AOV94584.1	GH0	glycogen debranching enzyme (alpha-1,6-glucosidase)	645	0
LC1Nh_0795	AHB42465.1	GH1	β-glucosidase; β-galactosidase; β-mannosidase	306	2.00E-98
LC1Nh_0129	PIZ00265.1	GH15	Glucoamylase; glucodextranase; α,α-trehalase	347	8.00E-115
LC1Nh_0131	AOV94591.1	GH57	α-amylase; α-galactosidase; amylopullulanase	715	0
LC1Hm_0354	ACV46754.1	CBM5	Cellulose-binding domain family V	1739	0
LC1Hm_2080	ACV48841.1	CBM5	Cellulose-binding domain family V	752	0
LC1Hm_0070	ACV47013.1	CBM5, GH18	Cellulose-binding domain family V, chitinase	734	0
LC1Hm_0809	ACV47620.1	CBM5, GH20	Cellulose-binding domain family V, β-hexosaminidase; lacto-N-biosidase	1405	0
LC1Hm_0822	ACV47608.1	CBM5, GH18	Cellulose-binding domain family V, chitinase	786	0
LC1Hm_2270	ACV49026.1	CBM5, GH18	Cellulose-binding domain family V, chitinase	976	0
LC1Hm_2271	ACV49027.1	CBM5, GH18	Cellulose-binding domain family V, chitinase	677	0
LC1Hm_2272	ACV49028.1	CBM5, GH18	Cellulose-binding domain family V, chitinase	1088	0
LC1Hm_2273	ACV49029.1	CBM5, GH18	Cellulose-binding domain family V, chitinase	867	0
LC1Hm_1158	BAC76692.1	GH18	Chitinase; lysozyme; endo-β-N-acetylglucosaminidase	271	1.00E-84
LC1Hm_0425	AEH39070.1	GH2	β-galactosidase; β-mannosidase; β-glucuronidase	1531	0
LC1Hm_1176	ADB63454.1	GH2	β-galactosidase; β-mannosidase; β-glucuronidase	952	0
LC1Hm_2335	AAV46707.1	GH2	β-galactosidase; β-mannosidase; β-glucuronidase	663	0
LC1Hm_4132	ACV49329.1	GH2	β-galactosidase; β-mannosidase; β-glucuronidase	1904	0
LC1Hm_1210	AGB37354.1	GH3	β-glucosidase; xylan 1,4-β-xylosidase; β-glucosylceramidase	722	0
LC1Hm_2524	ACV49279.1	GH3	β-glucosidase; xylan 1,4-β-xylosidase; β-glucosylceramidase	974	0
LC1Hm_2617	AKU08087.1	GH3	β-glucosidase; xylan 1,4-β-xylosidase; β-glucosylceramidase	982	0
LC1Hm_4049	AEH38218.1	GH4	Maltose-6-phosphate glucosidase; α-glucosidase; α-galactosidase	667	0
LC1Hm_0270	ACV46974.1	GH9	Endoglucanase; endo-β-1,3(4)-glucanase / lichenase-laminarinase; β-glucosidase	1001	0
LC1Hm_1562	ACV48372.1	GH13_31	α-amylase; pullulanase; cyclomaltodextrin glucanotransferase	1049	0
LC1Hm_2614	ACV46202.1	GH13_31	α-amylase; pullulanase; cyclomaltodextrin glucanotransferase	1036	0
LC1Hm_2615	ACV46203.1	GH13_31	α-amylase; pullulanase; cyclomaltodextrin glucanotransferase	1047	0
LC1Hm_1157	BAZ30250.1	GH29	α-L-fucosidase; α-1,3/1,4-L-fucosidase	289	7.00E-90
LC1Hm_1173	BAZ54052.1	GH29	α-L-fucosidase; α-1,3/1,4-L-fucosidase	347	3.00E-112
LC1Hm_1261	ACV48052.1	GH32	Invertase; endo-inulinase; β-2,6-fructan 6-levanbiohydrolase	1100	0
LC1Hm_1177	AGF92912.1	GH38	α-mannosidase; mannosyl-oligosaccharide α-1,2-mannosidase	916	0
LC1Hm_1260	ACV48051.1	GH68	Levansucrase; β-fructofuranosidase; inulosucrase	830	0
LC1Hm_4127	ACV49500.1	GH88	d-4,5-unsaturated β-glucuronyl hydrolase	764	0

Table S9. Molecular probes and conditions used in the current study for CARD-FISH (fluorescence *in-situ* hybridization and catalyzed reporter deposition).

Probes	Sequences (5'-3')	Formamide, (%, v/v)	Hybridization temperature (°C)	Washing temperature (°C)
Arch915[a]	GTGCTCCCCCGCCAATTCCT	20	46	48
Hmb_172[b]	CAAGCGTGAATGCGATAT	20	46	48
Narc_1214[c]	CCGCGTGTATCCCAGAGC	20	46	48

[a] (86)
[b] This study
[c] (27)

Table S10. Molecular probes designed and used in the current study for taxon-specific- and for quantitative PCR.

Organism	Primer	5' → 3'	Approach
Euryarchaea	Arch_20F	TTCCGGTTGATCCTGCCGG	Taxon-specific PCR
	Arch921R	TCCGGCGTTGAGTCCAATT	
Ca. Nanohalobium constans LC1Nh	Arc_20FM	TTCYGGTTGATCCTGCCRG	
	Nh_1179R	CGCGTGTATCCCAGAGCA	
Halomicrobium sp. LC1Hm	Hm_152F	GGATATCGCATTCACGCTTG	
	Arch921R	TCCGGCGTTGAGTCCAATT	
Ca. Nanohalobium constans LC1Nh	Nh_1014F	CGTGAGGTGTCCGGTTAAGT	Quantitative PCR
	Nh_1130R	GCTCCTTCCCCTGGTTTATC	
Halomicrobium sp. LC1Hm	Hm_0409	TTCTCGACCGTAAGGTGGTC	
	Hm_0527	CAAGCTACGGACGCTTTAGG	

Supplementary Figures

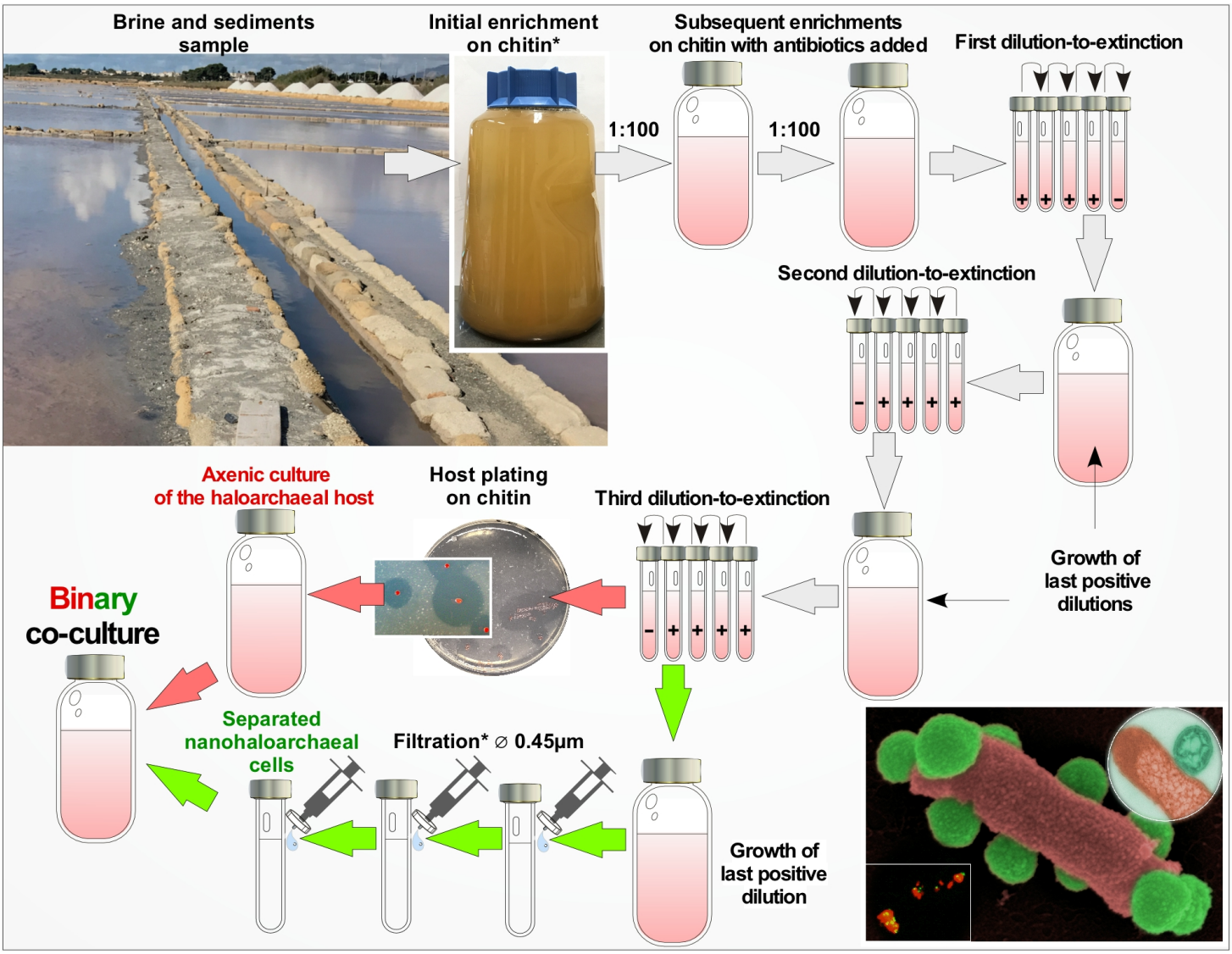

Figure S1. The crystallizer pond *Vasca* #27 (on the right) at *Saline delle Lagune* solar saltwork (Sicily) where sampling was carried out, and schematic of the laboratory-based strategy applied to obtain a binary co-culture *Ca.* Nanohalobium constans LC1Nh and its chitinotrophic host *Halomicrobium* sp. LC1Hm. We isolated the host by obtaining an axenic culture (red arrows), and isolated the nanohaloarchaeal cells using serial filtration (green arrows). Further details of cultivation- and dilution-to-extinction experiments are given in Material and Methods- and Results sections of the main text.

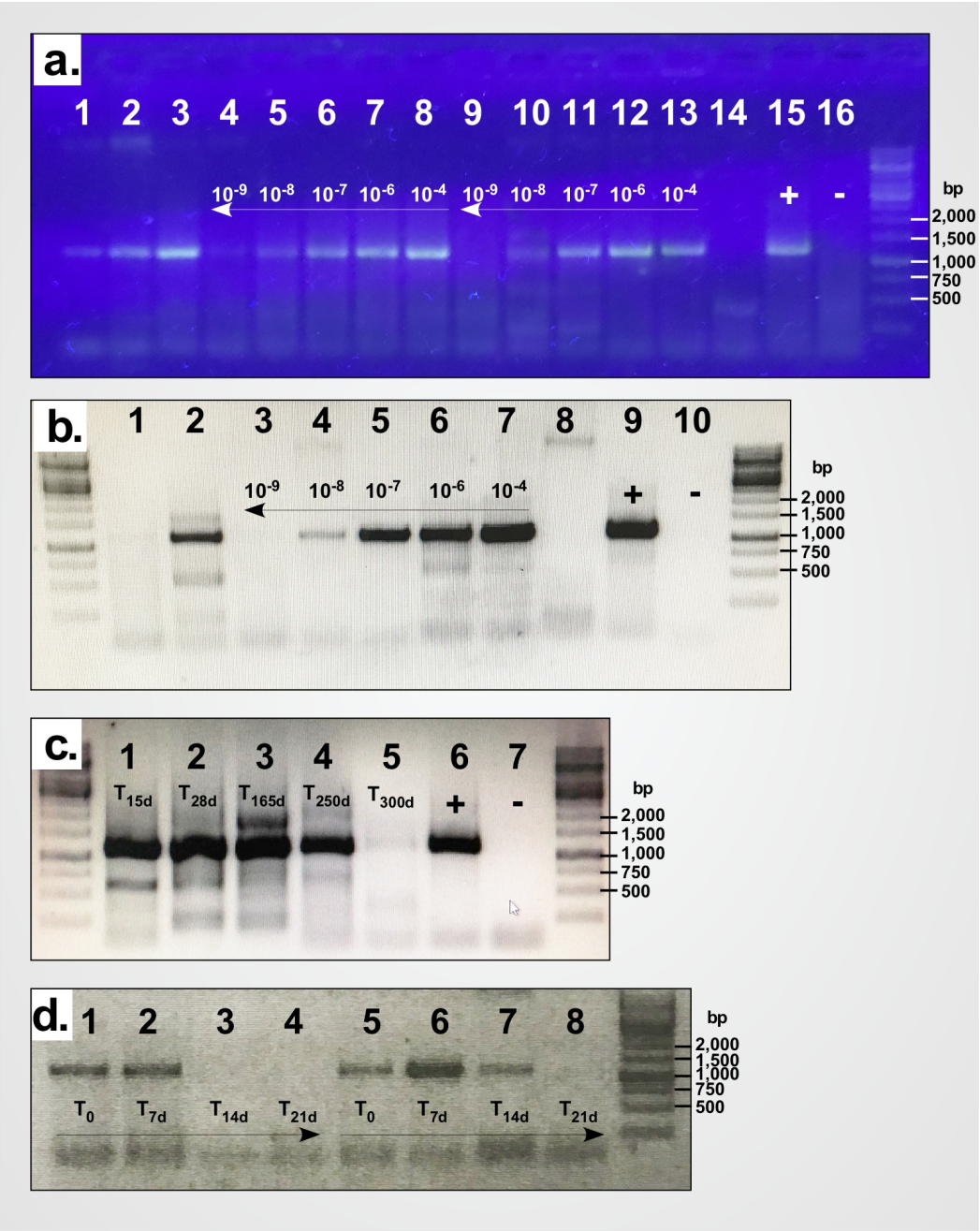

Figure S2. Taxon-specific PCR amplifications used to monitor:
(**a**) the initial enrichment on chitin (line 1); subsequent enrichments on chitin with antibiotics (lines 2 and 3); first dilution-to-extinction assay (lines 4-8), second dilution-to-extinction assay (lines 9-13); host colony grown on LC agar supplemented with chitin (line 14); positive control (16S rRNA gene of *Ca*. Nanohalobium constans LC1Nh, inserted in the pTA plasmid, [line 15]) and negative control (no DNA template, [line 16]); (**b**) the third round of 0.45-µm filtration; Euryarchaea- (line 1) and nanohaloarchaea-specific (line 2) 16S rDNA PCR-amplification; nanohaloarchaea-specific 16S rDNA PCR-amplification of the third dilution-to-extinction assay (lines 3-7) and the axenic culture of the *Halomicrobium* sp. LC1Hm host (line 8), positive control (16S rRNA gene of *Ca*. Nanohalobium constans LC1Nh, inserted in the pTA plasmid, [line 9]) and negative control (no DNA template, [line 10]);
(**c**) the binary *Ca*. Nanohalobium constans LC1Nh : Halomicrobium sp. LC1Hm co-culture stored at room temperature [20-25°C] in the dark for at least 300 days as confirmed by nanohaloarchaea-specific primers (lines 1-5); positive control (16S rRNA gene of *Ca*. Nanohalobium constans LC1Nh, inserted in the pTA plasmid, [line 6]) and negative control (no DNA template, [line 7]); and
(**d**) the nanohaloarchaeal cell-suspension, separated from the host by flitration and incubated during 3-weeks in the lower chamber of Millipore® Stericup® filtration system. Experiment was done in duplicate (lines 1-4 and lines 5-8); see Material and Methods for further details.

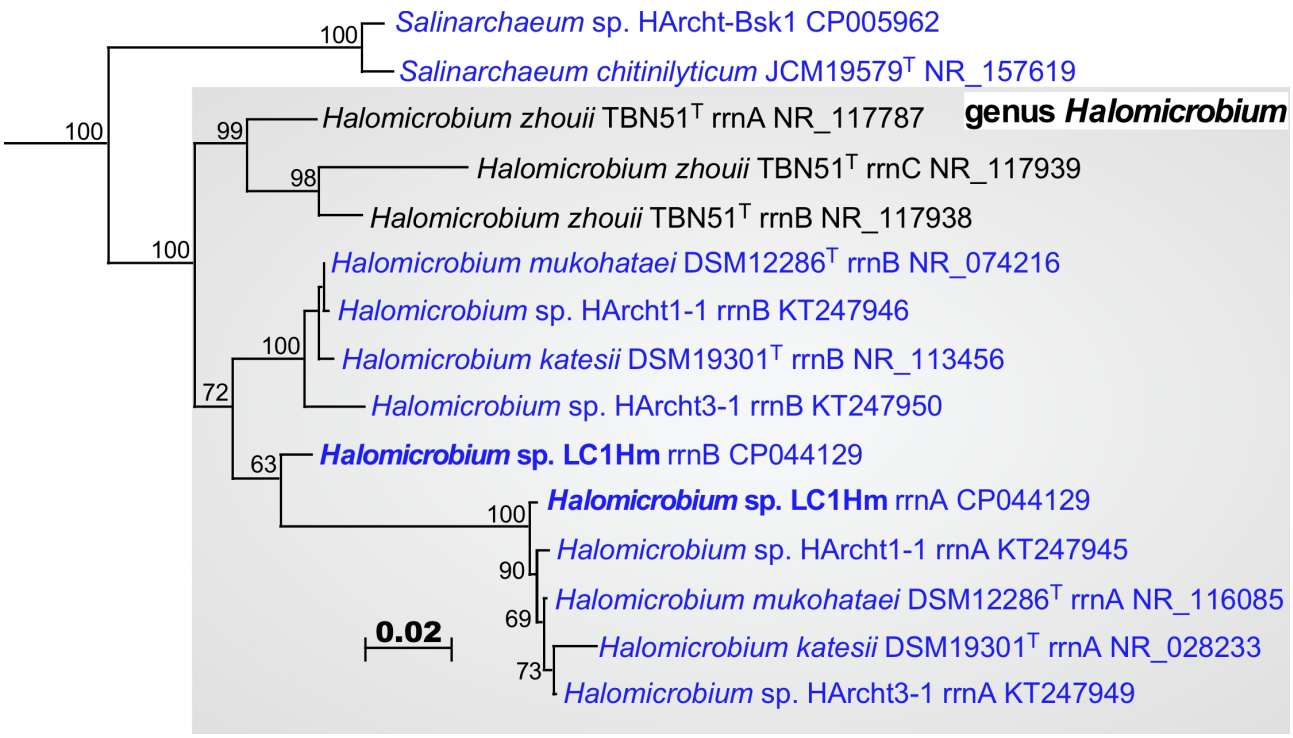

Figure S3. Maximum-likelihood 16S rRNA gene sequence-based phylogenetic tree, showing the position of strain LC1Hm (in bold) within the genus *Halomicrobium*. The strains with experimentally confirmed ability to grow on chitin as the single source of carbon and energy are indicated in blue. Branch lengths (see scale bar) correspond to the number of substitutions per site. All positions with less than 95% site coverage were eliminated. In total, 1435 positions were used in the tree, and numbers at nodes indicate bootstrap values of 1000 repetitions. The tree was rooted with two chitinolytic *Salinarchaeum* strains, and *Natronomonas pharaonis* JCM 8858^T (AB663432) was used as the outgroup.

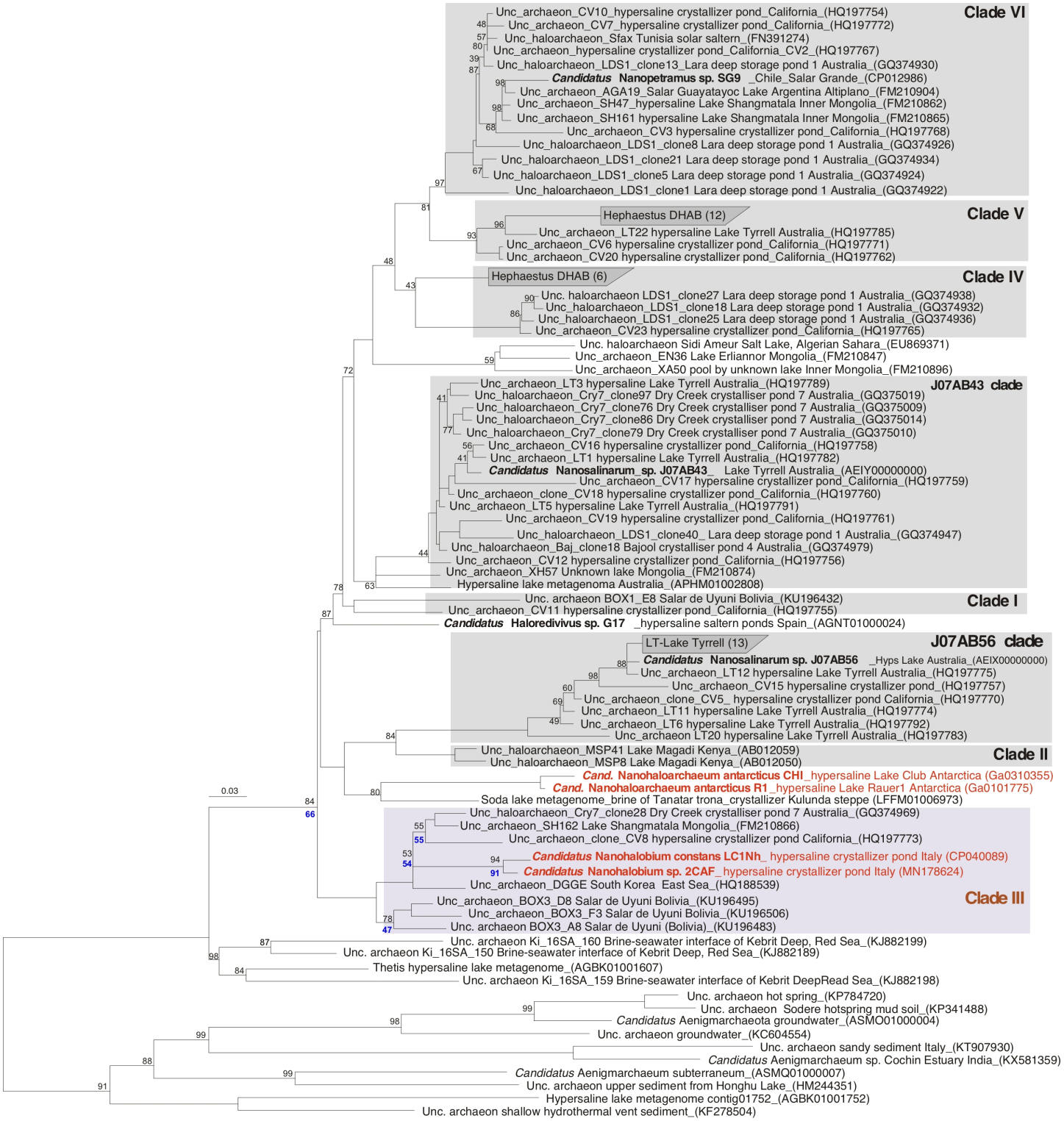

Figure S4. Phylogenetic position of *Ca.* Nanohalobium constans LC1Nh within the Nanohaloarchaeota. The tree was constructed based on alignment of 16S rRNA gene sequences to the SILVA Release 132SSURef NR99 using phylogenetic inference under maximum parsimony criteria within the ARB software environment. The tree was additionally inferred in the maximum likelihood framework using the MEGA v.6.0 software (87) and the bootstrap values supporting the positioning of *Ca*. Nanohalobium constans within the Clade III are shown in bold below the nodes. Ten 16S rRNA gene sequences of Aenigmarchaeota were used for out-grouping. Values less than 40% were omitted, and numbers at nodes represent bootstrap values (1000 replications of the original dataset). Scale bar represents the average number of substitutions per site. Cultivated nanohaloarchaea are highlighted in red.

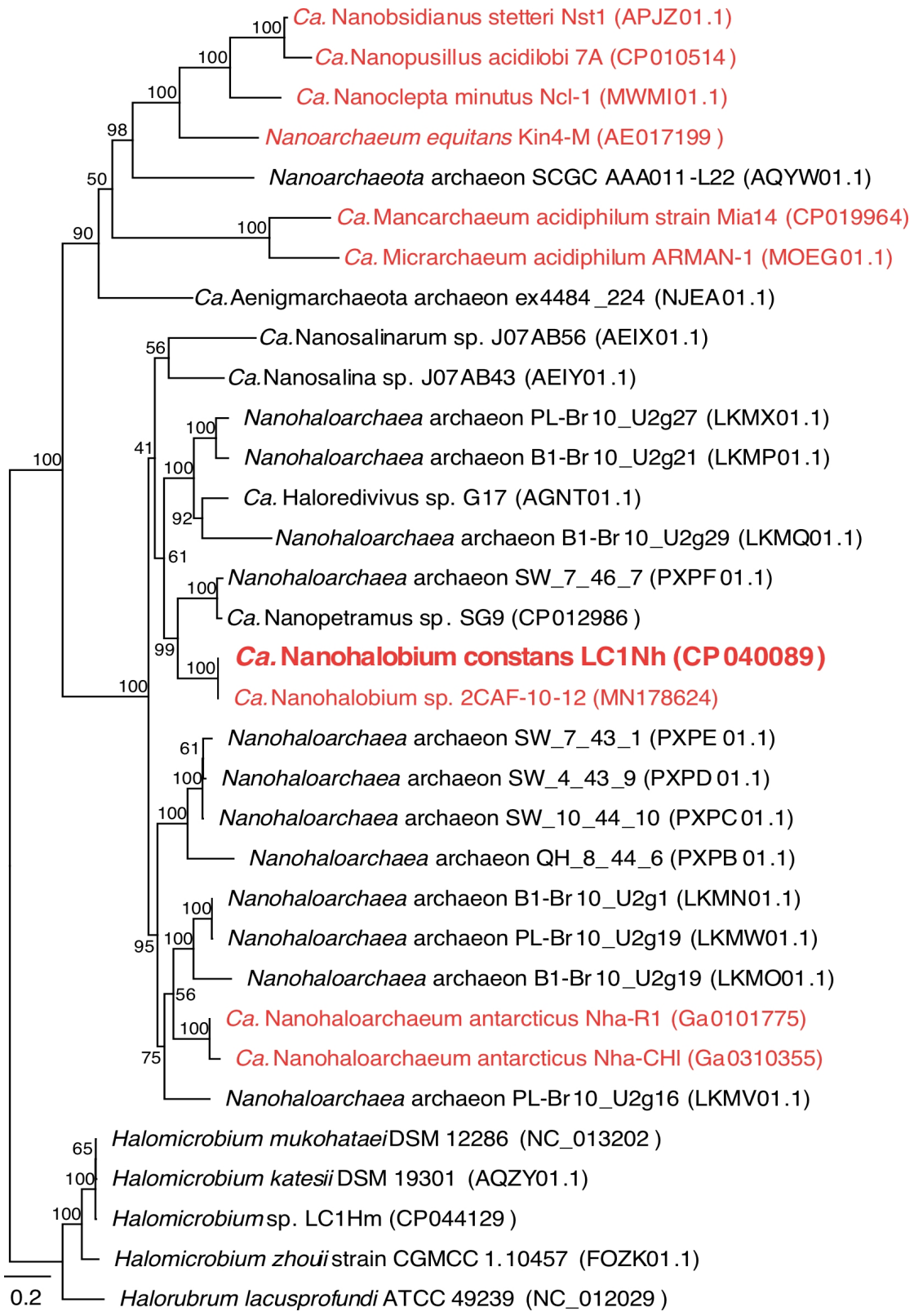

Figure S5. Maximum likelihood phylogeny of selected DPANN archaea based on the concatenated alignment of 11 ribosomal proteins. The amino-acid sequences were aligned by Clustal W 2.1 program with the BLOSUM (BLOcks SUbstitution Matrix) and the phylogeny was inferred by PhyML 3.0 plugin software inside Geneious 7.1 with the BLOSUM62 substitution model and 1,000 bootstrap replicates. Type species of *Halomicrobium*, together with the host *Halomicrobium* sp. LC1Hm and *Halorubrum lacusprofundi* ATCC 49239 were used as outgroup. Bootstrap support values (if >40) are indicated for selected groups at the tree's nodes. Cultured ectosymbiotic DPANN archaea are highlighted in red. The scale bar represents the average number of substitutions per site.

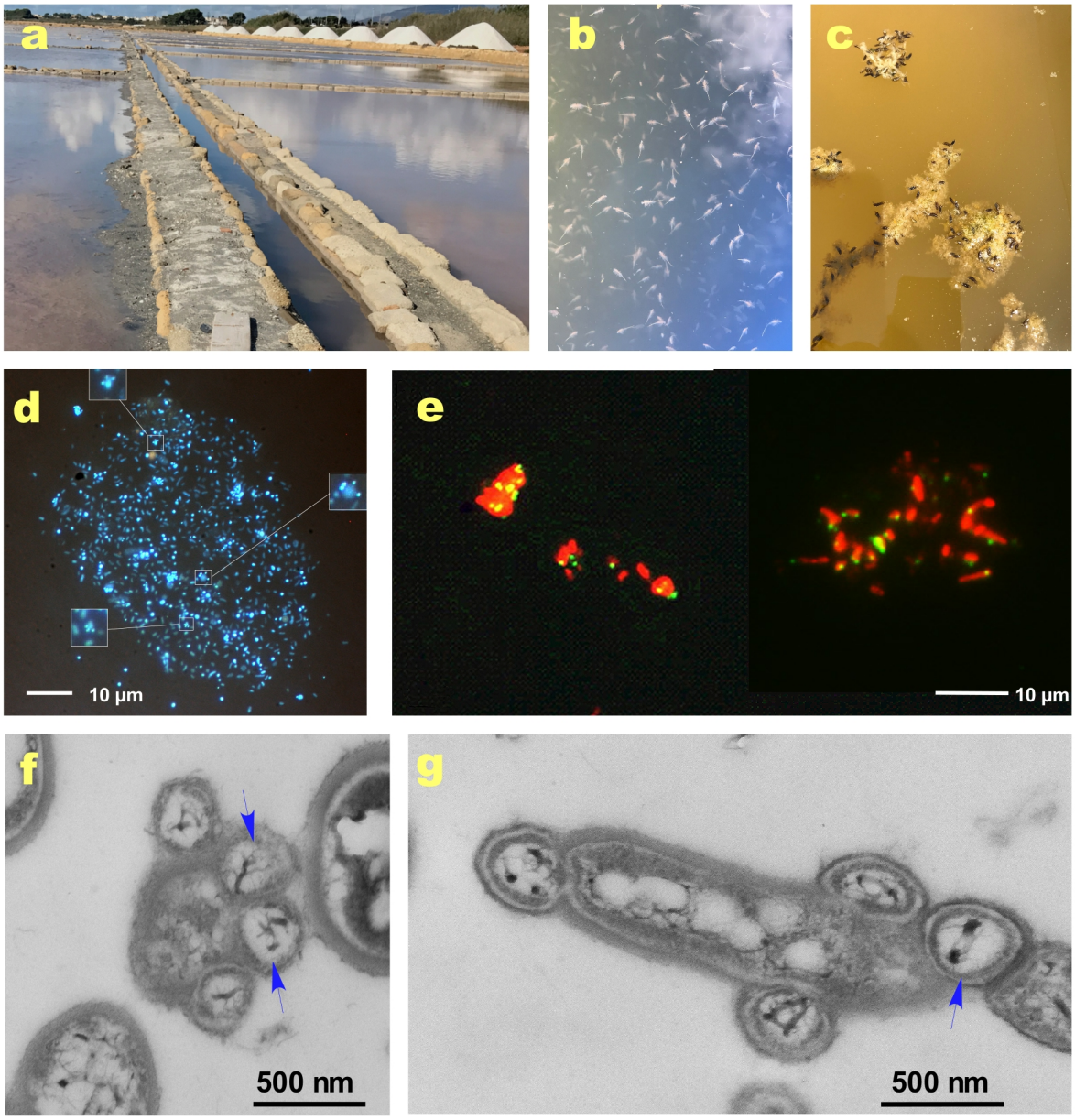

Figure S6. Site of environmental origin of *Ca.* Nanohalobium constans LC1Nh and *Halomicrobium* sp. LC1Hm and fluorescent and electron micrographs of *Ca.* Nanohalobium constans LC1Nh in co-culture with *Halomicrobium* sp. LC1Hm. Images show *Saline della Laguna* solar saltern system (37°51'48.70"N; 12°29'02.74"E), one of the ancient sites of salt production near Trapani (Sicily, Italy) (**a**); brine shrimps *Artemia* **(b**) within the brine, and brine flies (likely of the family *Ephydridea*) on the brine surface (**c)**, observed in different ponds of the system at the time of sampling; DAPI (**d**) and CARD-FISH (**e**) staining shows tiny coccoidal nanohaloarchaeal cells adhering to the host haloarchaeal cells (tyramide Alexa488 and Alexa594 were used for nanohaloarchaeal and haloarchaeal cells, respectively); transmission electron microscopy (TEM) images showing the close interaction of some *Ca.* Nanohalobium constans LC1Nh cells (chitoblue arrows) attached to more than one *Halomicrobium* sp. LC1Hm host cells (**f, g**).

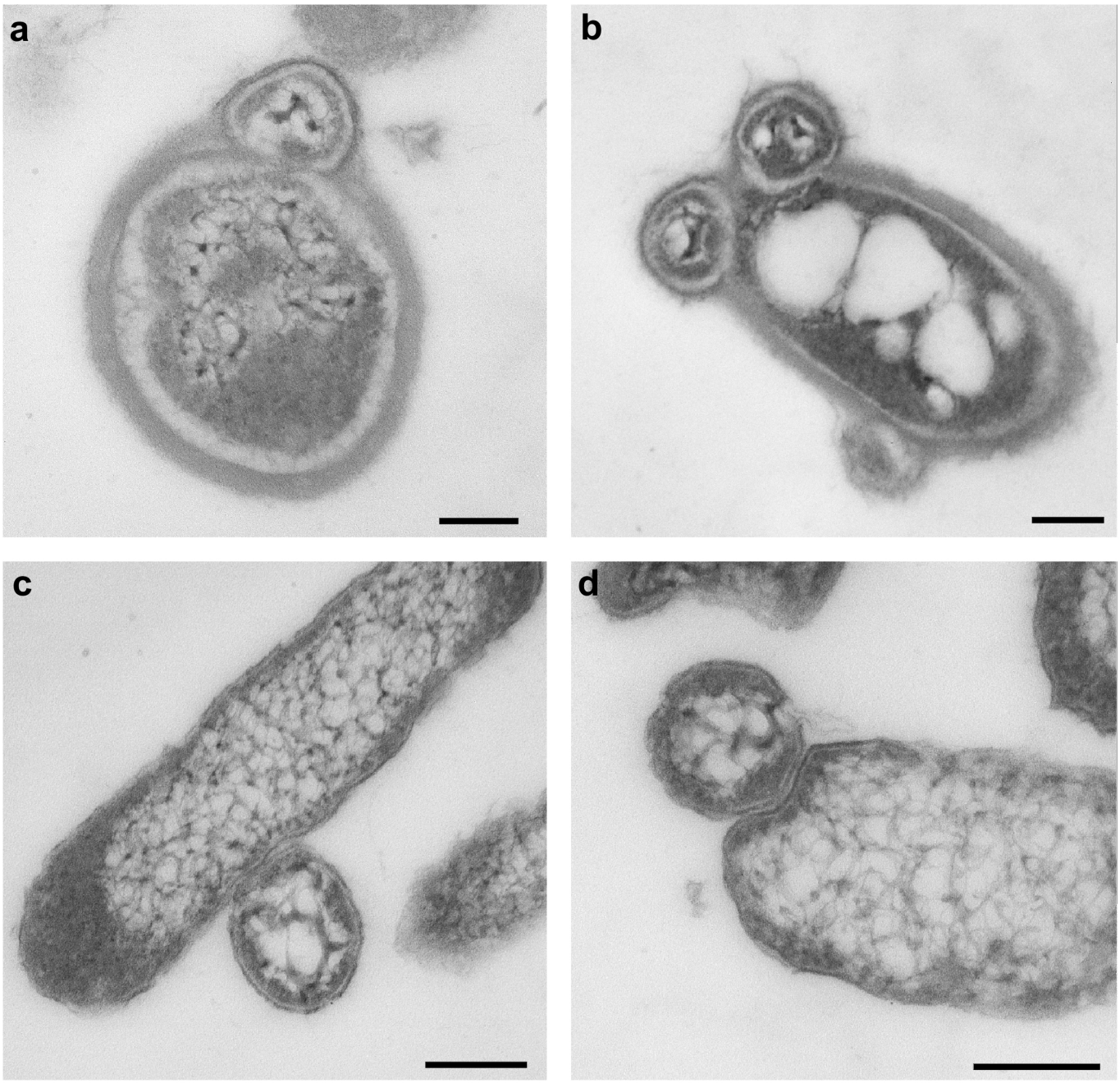

Figure S7. Transmission electron microscopy images reveal intimate contact between, and apparent fusion of, cells of *Ca.* Nanohalobium constans LC1Nh and the host *Halomicrobium* sp. LC1Hm. When grown on chitin (**a, b**), cells of *Halomicrobium* sp. LC1Hm have a thick, electron-dense external layer which is not observed during growth on the monosaccharide N-acetylglucosamine (**c**) or the disaccharide cellobiose (**d**). Scale bars represent 200 nm.

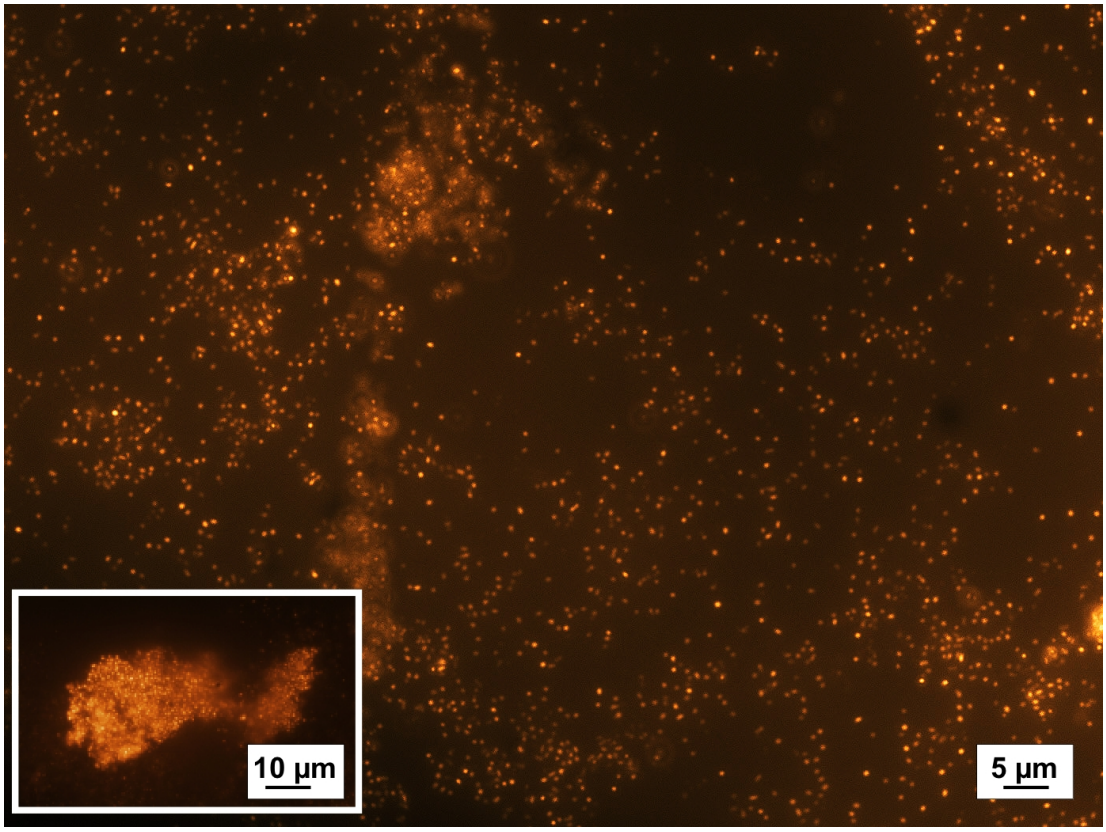

Figure S8. Fluorescence microscopy of chitinotrophic *Halomicrobium* sp. LC1Hm cells following staining with Nile Blue. Cells colonizing amorphous chitin and forming microcolonies shown in insert.

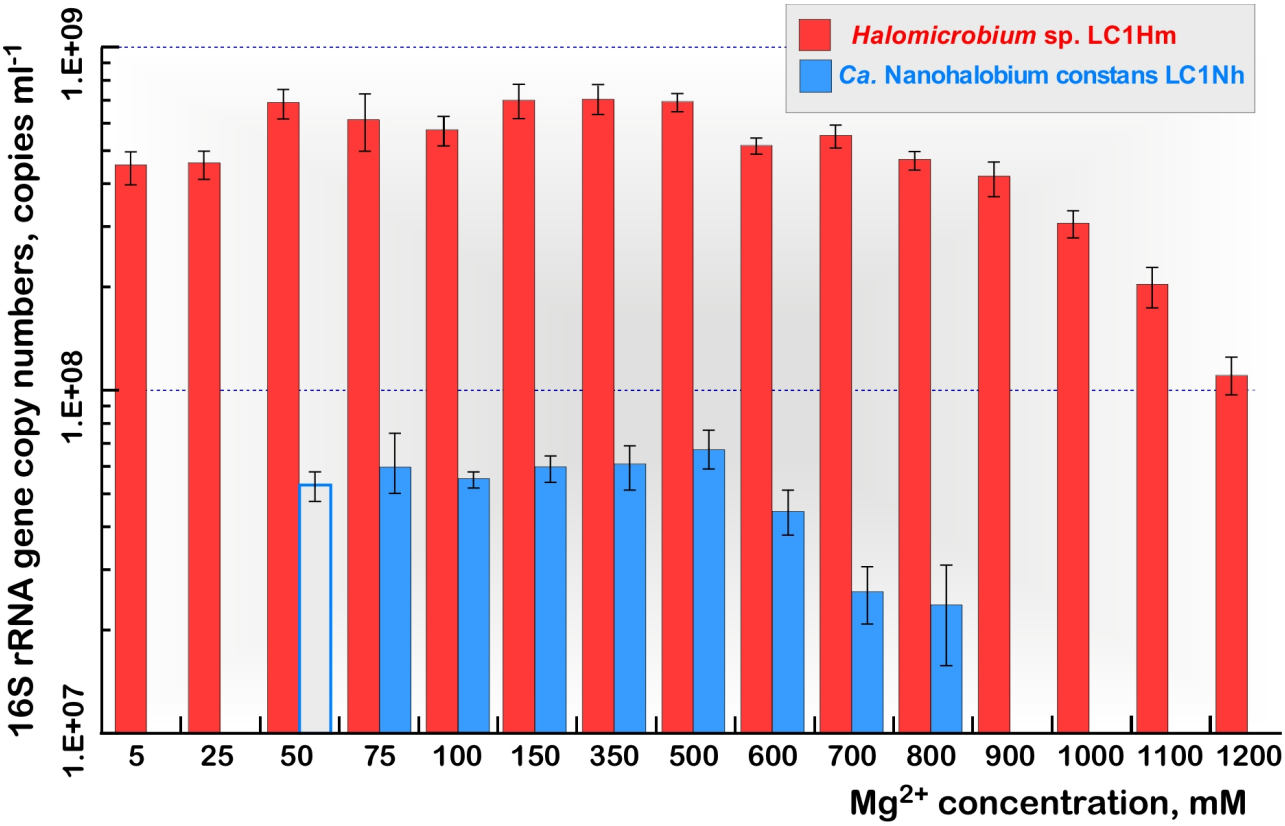

Figure S9. Magnesium tolerance and magnesium dependence of the extremely halophilic co-culture of *Halomicrobium* sp. LC1Hm and *Ca.* Nanohalobium constans LC1Nh. Cell densities were measured by qPCR analysis after 10 days of cultivation under microaerophilic conditions at 40°C and pH 7.2 using chitin (5 g l^{-1}) as the sole carbon- and energy source. Total salinity of LC culture medium was maintained as 240 g l^{-1} by varying either sodium or magnesium salts (see Materials and Methods). Growth of *Ca.* Nanohalobium constans LC1Nh at 50 mmol Mg^{2}+ concentration, shown as empty column, was observed only in the first generation of the *Halomicrobium* sp. LC1Hm + *Ca.* Nanohalobium constans LC1Nh co-culture, since the inoculum was taken from an exponential-phase co-culture growing at 315 mM Mg^{2+}. The plotted values (means) and standard deviations were calculated from three culture replicates.

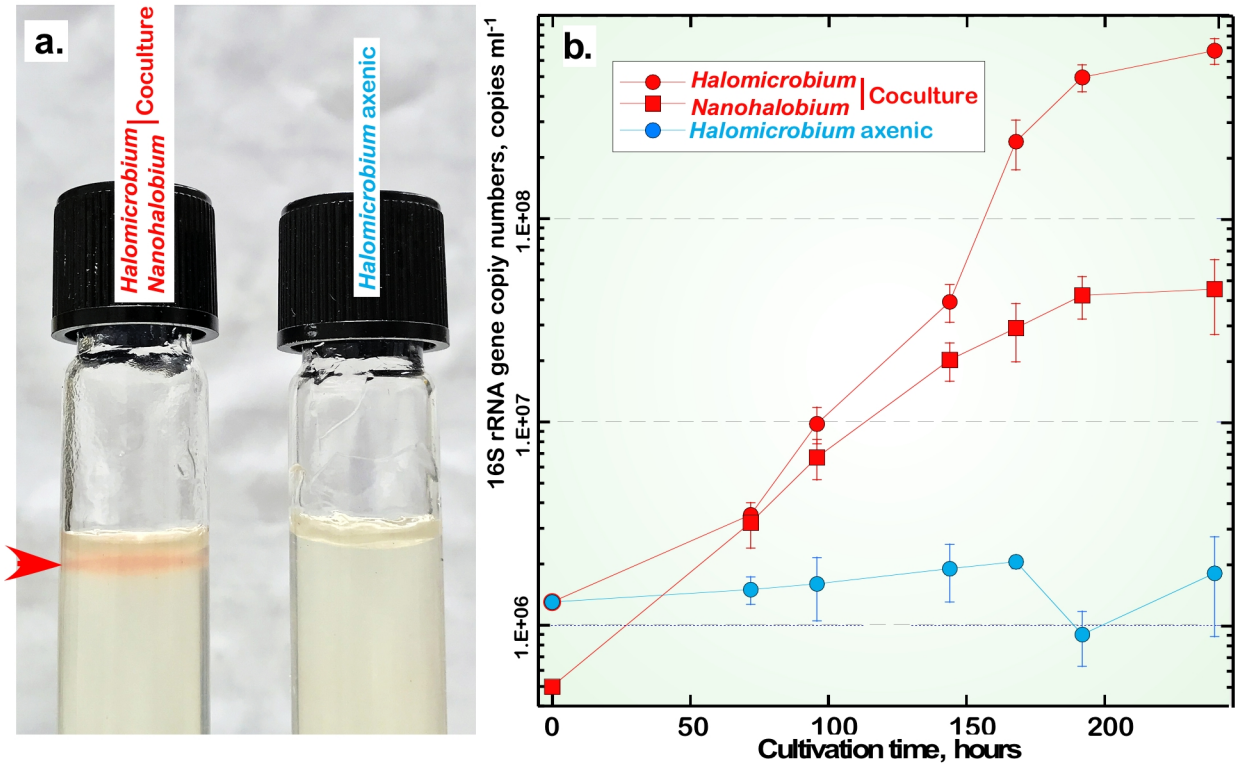

Figure S10. Indications of biomass for a *Ca.* Nanohalobium constans LC1Nh and *Halomicrobium* sp. LC1Hm co-culture, and an axenic culture of *Halomicrobium* sp. LC1Hm, on glycogen (2 g l^{-1}): (**a**) on LC agar (1.5% of agar, w/w); the red arrow indicates the zone of microaerophilic growth (a gradient of oxygen was achieved by placing of 1 ml of 2 mM Na_2S on the bottom of each Hungate tube); and (**b**) time course of each microbe in co-culture, and for *Halomicrobium* sp. LC1Hm in axenic culture; the plotted values (means) and standard deviations were calculated from three culture replicates.

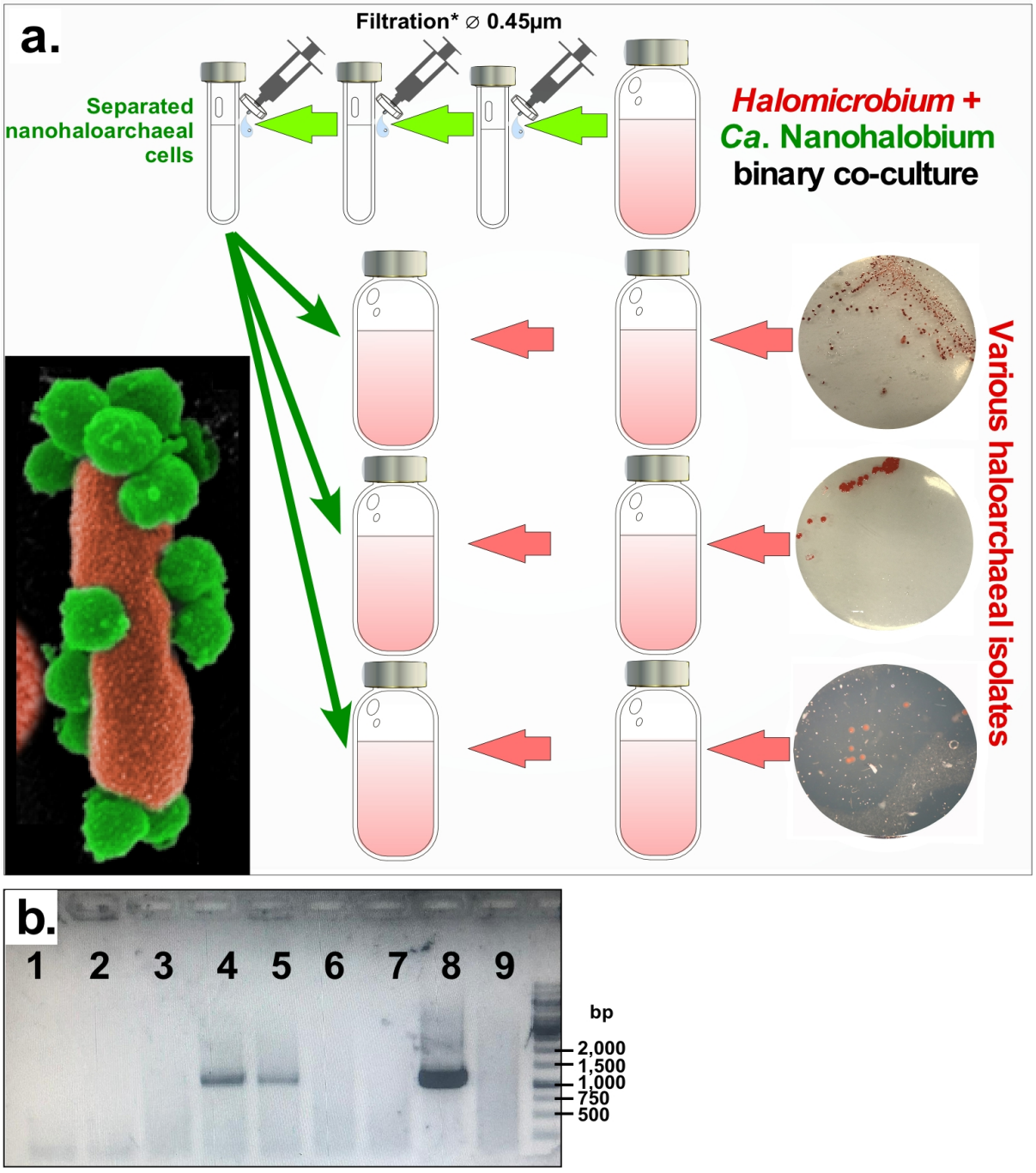

Figure S11. Schematic of the laboratory-based strategy to obtain the binary culture of *Ca.* Nanohalobium constans LC1Nh with various polysaccharidolytic and sugar-utilizing haloarchaea (**a**) and the nanohaloarchaeon-specific PCR amplifications applied to monitor the experiment (**b**). The DNA used for (**b**) was obtained from the following experiment: after one month of static incubation at 40°C in LC broth supplemented with 2 g l^{-1} of corresponding carbon sources (see Material and Mehods), the mixed cultures were checked on presence of nanohaloarchaea by taxon-specific 16S rDNA PCR amplification: chitinolytic *Natrinema* sp. HArcht2 (line 1); chitinolytic *Salinarchaeum* sp. HArcht-Bsk1 (line 2); three species of the genus *Halomicrobium*, cellulolytic *H. zhouii* JCM17095^T (line 3), chitinolytic *Halomicrobium* sp. HArcht3-1 (line 4) and the type species of the genus, chitinolytic *H. mukohataei* JCM9738^T (line 5); amylolytic *Halorhabdus thiamatea* SARL4B^T (line 6); sugar-utilizing haloarchaeon *Haloferax volcanii* DSM3757^T (line 7); reconstructed binary culture of *Halomicrobium* sp. LC1Hm + *Ca*. Nanohalobium constans LC1Nh (line 8); fraction of separated nanohaloarchaeal cells, incubated alone at 40°C in LC medium amended with 2 g l^{-1} of chitin (line 9).

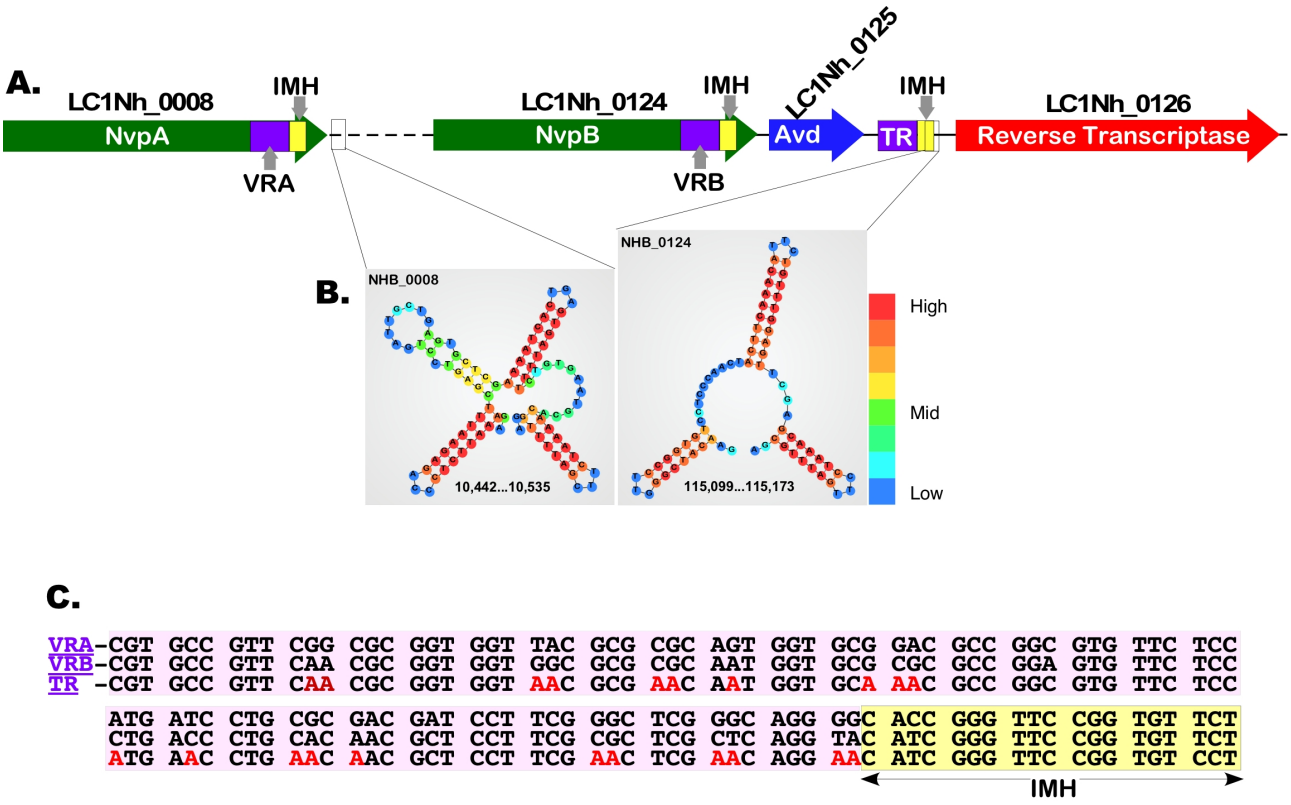

Figure S12. Diversity-generating retroelements (DGR) locus in *Ca.* Nanohalobium constans LC1Nh genome: (**A**) DGR locus consists of consists of the accessory variability determinant (Avd, LC1Nh_0125) and the error-prone reverse transcriptase (RT, LC1Nh_0126). Between the Avd and RT genes the LC1Nh genome contains a 95-bp long template region (TR). TR is similar to the variable regions (VRA and VRB) of two proteins. LC1Nh_0008 and LC1Nh_0123 (80% and 84%, respectively). At the 3'-end of the VRA, VRB and TR regions, the LC1Nh DGRs system composes of three identical 19-bp-long sequences, coined as initiation of mutagenic homing sequences (IMH); (**B**) two hairpin/cruciform structures downstream of the VRA and VRB were evident in the LC1Nh genome; and (**C**) alignment of VRA and VRB regions of *Ca*. Nanohalobium constans variable proteins NvpA and NvpB (LC1Nh_0008, 0123) with TR.

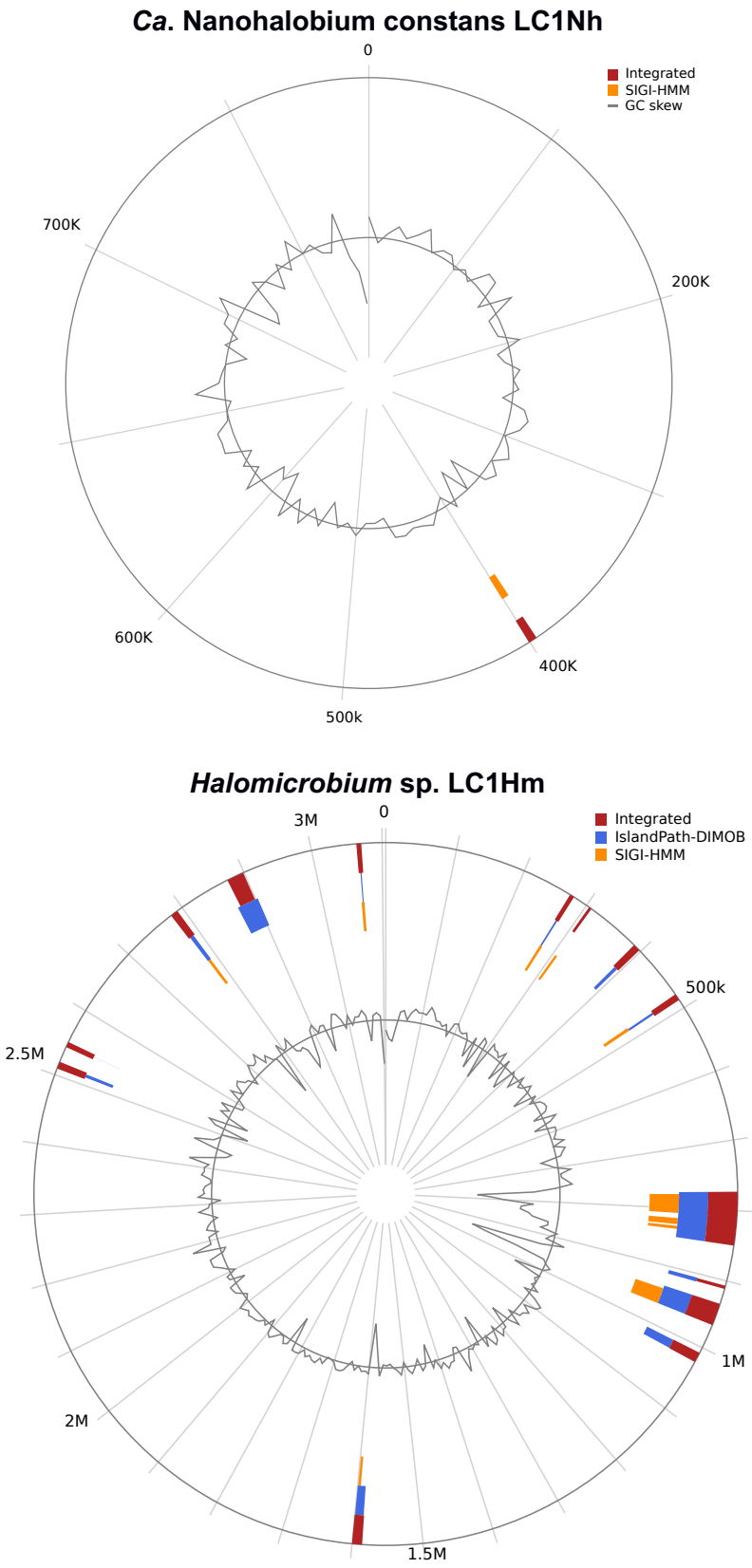

Figure S13. Genomic islands predicted with Islandviewer4. This tool uses four different prediction methods: IslandPath-DIMOB, SIGI-HMM, IslandPick and Islander. Only prediction methods with valid results are reported. The inner circle is the graphical representation of the GC skew predicted.

Figure S14. Two predicted incomplete prophages in the genome of *Ca*. Nanohalobium constans LC1Nh. The list of gene products is present in SI Extended Dataset S3d and S3e.

SI Extended Dataset S1 (separate file)
S1a – BlastKOALA annotations list of *Halomicrobium* sp. LC1Hm chromosome
S1b – BlastKOALA results classified in KEGG pathways for *Halomicrobium* sp. LC1Hm chromosome
S1c – BlastKOALA annotations list of *Halomicrobium* sp. LC1Hm plasmid pLC1HM-01
S1d – BlastKOALA results classified in KEGG pathways for *Halomicrobium* sp. LC1Hm plasmid pLC1HM-01
S1e – NetNGlyc prediction results list for *Ca*. Nanohalobium constans LC1Nh genome.

SI Extended Dataset S2 (separate file)
S2a – Proteome of *Ca*. Nanohalobium constans LC1Nh grown on chitin in co-culture with *Halomicrobium* sp. LC1Hm
S2b – Proteome of *Halomicrobium* sp. LC1Hm grown on chitin in co-culture with *Ca*. Nanohalobium constans LC1Nh.

SI Extended Dataset S3 (separate file)
S3a – Groups of similar proteins identified in *Ca*. Nanohalobium constans LC1Nh and *Halomicrobium* sp. LC1Hm genomes with ProteinOrtho
S3b – List of *Ca*. Nanohalobium constans LC1Nh genes ordered by GC content
S3c – Results of blastp and HGT detection for *Ca*. Nanohalobium constans LC1Nh non ribosomal genes with GC > 50%
S3d - Summary list of PHAST results for Candidatus Nanohaloarchaeota WGS projects registered in the NCBI (JGI) and *Halomicrobium* sp. LC1Hm
S3e – Detailed list of PHAST results for Candidatus Nanohaloarchaeota WGS projects registered in the NCBI (JGI) and *Halomicrobium* sp. LC1Hm
S3f – List of Genomic Island (Gi) detected in *Ca*. Nanohalobium constans LC1Nh and *Halomicrobium* sp. LC1Hm genomes with IslandViewer4.

SI References

1. M. Y. Galperin, K. S. Makarova, Y. I. Wolf, E. V. Koonin, Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* **43**, D261–D269 (2015).
2. K. S. Makarova, A. V. Sorokin, P. S. Novichkov, Y. I. Wolf, E. V. Koonin, Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biol. Direct.* **2**, 33 (2007).
3. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Archaeal clusters of orthologus genes (arCOGs): an update and application for analysis of shared features between *Thermococcales*, *Methanococcales* and *Methanobacteriales*. *Life (Basel)* **5**, 818–840 (2015).
4. B. J. Tindall *et al.,* Complete genome sequence of *Halomicrobium mukohataei* type strain (arg-2). *Stand. Genomic Sci.* **1**, 270−277 (2009).
5. J. Hou *et al.*, Characterization of genes for chitin catabolism in *Haloferax mediterranei*. *Appl. Microbiol. Biotechnol.* **98**, 1185–1194 (2014).
6. D. Y. Sorokin, S. V. Toshchakov, T. V. Kolganova, I. V. Kublanov, Halo(natrono)archaea isolated from hypersaline lakes utilize cellulose and chitin as growth substrates. *Front. Microbiol.* **6**, 942 (2015).
7. J. Quillaguamán, H. Guzmán, D. Van-Thuoc, R. Hatti-Kaul, Synthesis and production of polyhydroxyalkanoates by halophiles: Current potential and future prospects. *Appl. Microbiol. Biotechnol.* **85**, 1687–1696 (2010).
8. A. Koller, Polyhydroxyalkanoate biosynthesis at the edge of water activity – haloarchaea as biopolyester factories. *Bioengineering* **6**, 34 (2019).
9. N. Dombrowski, J. H. Lee, T. A. Williams, P. Offre, A. Spang, Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol. Lett.* **366**, fnz008 (2019).
10. J. N. Hamm *et al.*, Unexpected host dependency of Antarctic Nanohaloarchaeota. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14661–14670 (2019).
11. H. Huber *et al.*, A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63–67 (2002).
12. U. Jahn *et al.*, *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two archaea. *J Bacteriol.* **190**, 1743–1750 (2008).
13. O. V. Golyshina *et al.*, 'ARMAN' archaea depend on association with euryarchaeal host in culture and in situ. *Nat. Commun.* **8**, 60 (2017).
14. S. Krause *et al.*, Characterisation of a stable laboratory co-culture of acidophilic nanoorganisms. *Sci. Rep.* **7**, 3289 (2017).
15. J. K. Jarett *et al.*, Single-cell genomics of co-sorted Nanoarchaeota suggests novel putative host associations and diversification of proteins involved in symbiosis. *Microbiome* **6**, 161 (2018).
16. M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).

17. S. V. Albers, K. F. Jarrell, The archaellum: an update on the unique archaeal motility structure. *Trends Microbiol.* **26**, 351–362 (2018).

18. C. J. Castelle *et al.*, Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690–701 (2015).

19. C. J. Castelle, C. T. Brown, K. Anantharaman, A. J. Probst, R. H. Huang, J. F. Banfield, Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat. Rev. Microbiol.* **16**, 629–645 (2018).

20. B. Medhekar, J. F. Miller, Diversity-generating retroelements. *Curr. Opin. Microbiol.* **10**, 388–395 (2007).

21. B. G. Paul *et al.*, Retroelement-guided protein diversification abounds in vast lineages of Bacteria and Archaea. *Nat. Microbiol.* **2**, 17045 (2017).

22. S. Handa, B. G. Paul, J. F. Miller, D. L. Valentine, P. Ghosh, Conservation of the C-type lectin fold for accommodating massive sequence variation in archaeal diversity-generating retroelements. *BMC Struct. Biol.* **16**, 13 (2016).

23. H. Guo *et al.*, Target site recognition by a diversity-generating retroelement. *PLoS Genet.* **7**, e1002414 (2011).

24. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

25. J. Le Coq, P. Ghosh, Conservation of the C-type lectin fold for massive sequence variation in a Treponema diversity-generating retroelement. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14649–14653 (2011).

26. B. Geueke, B. Riebel, W. Hummel, NADH oxidase from Lactobacillus brevis: a new catalyst for the regeneration of NAD. *Enzyme Microb. Technol.* **32**, 205–211 (2003).

27. K. Ito, K. Inaba, The disulfide bond formation (Dsb) system. *Curr. Opin. Struct. Biol.* **18**, 450–458 (2008).

28. P. Narasingarao *et al.*, De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J.* **6**, 81−93 (2012).

29. R. Ghai *et al.*, New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.* **1**, 739–751, (2011).

30. K. M. Finstad *et al.*, Microbial community structure and the persistence of cyanobacterial populations in salt crusts of the hyperarid Atacama Desert from genome-resolved metagenomics. *Front. Microbiol.* **8**, 1435 (2017).

31. F. Gonzalez-Ordenes *et al.*, ADP-Dependent Kinases From the Archaeal Order Methanosarcinales Adapt to Salt by a Non-canonical Evolutionarily Conserved Strategy. *Front. Microbiol.* **9**, 1305 (2018).

32. C. Bräsen, D. Esser, B. Rauch, B. Siebers, Carbohydrate metabolism in Archaea: current insights into unusual enzymes and pathways and their regulation. *Microbiol. Mol. Biol. Rev.* **78**, 89–175 (2014).

33. H. Imanaka, A. Yamatsu, T. Fukui, H. Atomi, T. Imanaka, Phosphoenolpyruvate synthase plays an essential role for glycolysis in the modified Embden-Meyerhof pathway in *Thermococcus kodakarensis*. *Mol. Microbiol.* **6**, 898–909 (2006).

34. M. Falb *et al.*, Metabolism of halophilic archaea. *Extremophiles* **12**,177–196 (2008).

35. B. Siebers, P. Schönheit, Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. *Curr. Opin. Microbiol.* **8**, 695−705 (2005).

36. M. Lechner *et al.*, Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**, 124 (2011).

37. M. Ravenhall, N. Škunca, F. Lassalle, C. Dessimoz, Inferring horizontal gene transfer. *PLoS Comput. Biol.* **11**, e1004095 (2015).

38. A. Boc, H. Philippe, V. Makarenkov, Inferring and validating horizontal gene transfer events using bipartition dissimilarity. *Syst. Biol.* **59**, 195–211 (2010).

39. A. Boc, A., A. B. Diallo, V. Makarenkov, T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res.* **40**, W573–W579 (2012).

40. C. Bertelli *et al.*, IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* **45**, W30–W35 (2017).

41. Y. Zhou, Y. Liang, K. H. Lynch, J. J. Dennis, D. S. Wishart, PHAST: a fast phage search tool. *Nucleic Acids Res.* **39**, W347–W352. (2011).

42. D. Arndt *et al.*, PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).

43. M. Martínez-García, F. Santos, M. Moreno-Paz, V. Parro, J. Antòn, Unveiling viral–host interactions within the 'microbial dark matter'. *Nat. Commun.* **5**, 4542 (2014).

44. M. M. Yakimov *et al.*, Microbial life in the Lake Medee, the largest deep-sea salt-saturated formation. *Sci. Rep.* **3**, 3554 (2013).

45. S. E. Dyksterhouse, J. P. Gray, R. P. Herwig, J. C. Lara, J. T. Staley, *Cycloclasticus pugetii* gen. nov., an aromatic hydrocarbon-degrading bacterium from marine sediments. *Int. J. Sys. Bacteriol.* **45**, 116–123 (1995).

46. N. Pfennig, K. D. Lippert, Über das Vitamin B12-Bedürfnis phototropher Schwefelbakterien. *Arch. Mikrobiol.* **55**, 245–256 (1966).

47. J. E. Hallsworth, N. A Magan, Rapid HPLC protocol for detection of polyols and trehalose. *J. Microbiol. Meth.* **1**, 7–13 (1997).

48. D. J. Lane, "16/23S rRNA sequencing" (John Wiley & Sons, New York, 1991).

49. E. F. De Long, Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5685–5689 (1992).

50. V. La Cono *et al.*, Unveiling microbial life in new deep-sea hypersaline Lake Thetis. Part I: Prokaryotes and environmental settings. *Environ. Microbiol.* **13**, 2250–2268 (2011).

51. J. G. Caporaso *et al.*, Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).

52. D. P. Herlemann *et al.*, Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* **5**, 1571–1579 (2011).

53. A. Klindworth *et al.*, Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, e1 (2013).

54. C. Quast *et al.*, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).

55. E. Pruesse, J. Peplies, F. O. Glöckner, SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829 (2012).

56. W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

57. C. Camacho *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

58. J. Pernthaler, F. O. Glockner, W. Schonhuber, R. I. Amann, "Fluorescence in situ hybridization (FISH) with rRNA-target oligonucleotide probes" (Academic Press, San Diego, 2001) Vol. 30.

59. A. G. Ostle, J. G. Holt, Nile Blue A as a fluorescent stain for poly-b-hydroxybutyrate. *Appl. Env. Microbiol.* **44**, 238–241 (1982).

60. A. Loy *et al.*, ProbeCheck—a central resource for evaluating oligonucleotide probe coverage and specificity. *Environ. Microbiol.* **10**, 2894–2898 (2008).

61. R. Rasmussen, "Rapid Cycle Real-Time PCR: Methods and Applications" (Springer Press, Berlin, Heidelberg, 2001).

62. R. Schmieder, R. Edwards, Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863−864 (2011).

63. T. Magoč, S. L. Salzberg, FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

64. R. R. Wick, L. M. Judd, C. L. Gorrie, K. E. Holt, Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **13**, e1005595 (2017).

65. A. L. Delcher, K. A. Bratke, E. C. Powers, S. L. Salzberg, Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673–679 (2007).

66. P. Rice, I. Longden, A. Bleasby, EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276−277 (2000).

67. V. Solovyev, A. Salamov, "Automatic Annotation of Microbial Genomes and Metagenomic Sequences" (Nova Science Publishers, Hauppauge, New York, 2011).

68. S. F. Altschul *et al.*, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

69. R. L. Tatusov, E. V. Koonin, D. J. Lipman, A genomic perspective on protein families. *Science* **278**, 631−637 (1997).

70. R. K. Aziz *et al.*, The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).

71. T. M. Lowe, P. P. Chan, tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **44**, W54–W57 (2016).

72. K. Lagesen *et al.*, RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).

73. V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, The Carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).

74. A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).

75. M. Krzywinski *et al.*, Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

76. P. A. Chaumeil, A. J. Mussig, P. Hugenholtz, D. H. Parks, GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, btz848 (Online ahead of print) (2019).

77. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

78. M. Anisimova, M. Gil, J. F. Dufayard, C. Dessimoz, O. Gascuel, Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst. Biol.* **60**, 685–699 (2011).

79. V. Lefort, J. E. Longueville, O. Gascuel, SMS: Smart Model Selection in PhyML. *Mol. Biol. Evol.* **34**, 2422–2424 (2017).

80. M. A. Larkin *et al.*, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).

81. S. Guindon *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

82. R. C. Valentine, B. M. Shapiro, E. R. Stadtman, Regulation of glutamine synthetase. XII Electron microscopy of the enzyme from *Escherichia coli*. *Biochem.* **7**, 2143−2152 (1968).

83. O. Langella *et al.*, X!TandemPipeline: a tool to manage sequence redundancy for protein inference and phosphosite identification. *J. Proteome Res.* **16**, 494–503 (2017).

84. Y. Ishihama *et al.*, Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Mol. Cell. Proteomics* **4**, 1265–1272 (2005).

85. J. Rappsilber, U. Ryder, A. I. Lamond, M. Mann, Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**, 1231–1245 (2002).

86. D. A. Stahl, R. Amann, “Development and application of nucleic acid probes” in Nucleic Acid Techniques in Bacterial Systematics, E. Stackebrandt, M. Goodfellow, Eds. (Chichester, UK: John Wiley and Sons Ltd, UK, 1991), pp. 205–248.

87. K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).