ORIGINAL ARTICLE (*Microbial Ecology*)

Taxonomic and functional metagenomic profiling of the microbial community in the anoxic sediment of a sub-saline shallow lake (Laguna de Carrizo – Central Spain)

Manuel Ferrer[1,§], María-Eugenia Guazzaroni[1,*], Michael Richter[2,*], Adela García-Salamanca[3,*], Pablo Yarza[2*], Ana Suárez-Suárez[2], Jennifer Solano[4], María Alcaide[1], Pieter van Dillewijn[3], M. Antonia Molina-Henares[3], Nieves López-Cortés[1], Yamal Al-Ramahi[1], Carmen Guerrero[5], Alejandro Acosta[3], Laura I. de Eugenio[6], Virginia Martínez[6], Silvia Marques[3], Fernando Rojo[7], Eduardo Santero[8], Olga Genilloud[9], Julian Pérez-Pérez[10], Ramón Rosselló-Móra[2] and Juan Luis Ramos[3]

[1]*CSIC, Institute of Catalysis, 28049 Madrid, Spain*
[2]*IMEDEA (CSIC-UIB), 07190 Esporles, Spain*
[3]*CSIC, Estación Experimental del Zaidín, 18008 Granada, Spain*
[4]*Bio-Illiberis R&D, 18100 Armilla, Granada, Spain*
[5]*Universidad Autónoma de Madrid, Facultad de Ciencias, 28049 Madrid, Spain*
[6]*CSIC, Centro de Investigaciones Biológicas, 28040 Madrid, Spain*
[7]*CSIC, Centro Nacional de Biotecnología, 28049 Madrid, Spain*
[8]*Universidad Pablo de Olavide, 41013 Sevilla, Spain*
[9]*Fundación Medina, 18100 Armilla, Granada, Spain*
[10]*Secugen, 28049 Madrid, Spain*

*These authors contributed equally to the work

[§]For correspondence. E-mail: mferrer@icp.csic.es; Tel (+34) 91 585 4872; Fax: (+34) 91 585 4760.

Running title: Metagenomic analysis of an anoxic sub-saline sediment

Key words: Biodiversity - metagenome - microbe - anoxic - subterranean water

This Whole Genome Shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession number ADZX00000000. The 16S rRNA genes sequences have been deposited in DDBJ/EMBL/GenBank under the accession number HQ003464-HQ003710. The version described in this paper is the first one, ADZX01000000.

Abstract

The phylogenetic and functional structure of the microbial community residing in a Ca^{2+}-rich anoxic sediment of a sub-saline shallow lake (Laguna de Carrizo, initially operated as a gypsum ($CaSO_4$ x 2 H_2O) mine) was estimated by analyzing the diversity of 16S rRNA amplicons and a 3.1 Mb of consensus metagenome sequence. The lake has about half the salinity of seawater and possesses an unusual relative concentration of ions, with Ca^{2+} and SO_4^{2-} being dominant. The 16S rRNA sequences revealed a diverse community with about 22% of the bacterial rRNAs being less than 94.5% similar to any rRNA currently deposited in GenBank. In addition to this, about 79% of the archaeal rRNAs genes were mostly related to uncultured *Euryarchaeota* of the CCA47 group, which are often associated with marine and oxygen-depleted sites. Sequence analysis of assembled genes revealed that 23% of the open reading frames of the metagenome library had no hits in the database. Among annotated genes, functions related to (thio) sulfate and (thio) sulfonate-reduction and iron-oxidation, sulfur-oxidation, denitrification, synthrophism and phototrophic sulfur metabolism were found as predominant. Phylogenetic and biochemical analyses indicate that the inherent physical-chemical characteristics of this habitat coupled with adaptation to anthropogenic activities have resulted in a highly efficient community for the assimilation of polysulfides, sulfoxides and organosulfonates together with nitro-, nitrile- and cyanide-substituted compounds. We discuss that the relevant microbial composition and metabolic capacities at Laguna de Carrizo, likely developed as an adaptation to thrive in the presence of moderate salinity conditions and potential toxic bio-molecules, contrast with the properties of previously known anoxic sediments of shallow lakes.

Introduction

Freshwater ecosystems, which account for circa 2.5% of the total volume of water available on our planet, are extremely different in composition and according to that distinct microbial communities have been established, as revealed by both cultivation and molecular-based and, more recently, by meta-genomic [10] approaches. Some major discoveries have been made in the last years on very peculiar microbial life adapted to the water column of many lakes; however, so far, only little is known and even less is understood about the microbial ecology and gene inventory of anoxic freshwater lake sediments.

Anoxic lake sediments from around the world, including those from saline and alkaline soda lakes [2, 72], hypersaline lakes [73], athalassohaline lakes [28], shallow suboxic-to-anoxic

freshwater ponds [5], sulfurous karstic lake [44], eutrophic lakes –including shallow [17, 48, 74, 76], sulfur-rich minerotrophic peatlands [31], warm monomictic and meso-eutrophic lakes [68, 69], freshwater tidal marshes [82], meromictic lakes [39, 42], as well as metal mining-impacted lakes [8, 13, 19, 59], have been studied, but most of this interest is centred on their phylogeny. Most of the communities were dominated by (un)culturable methane-producing archaea *Methanomicrobiales*, *Methanobacteriaceae* and *Methanosarcinales* and *Crenarchaeota* from uncultivable groups such as Miscellaneous Crenarchaeota group, Marine Group I, Marine Benthic Group B and C, Freshwater group, Group I3 and Rice Clusters IV and VI. *Crenarchaeota* represented the majority of the microbial population in mercury-contaminated freshwater stream [59] and sulfurous karstic lake sediments [44]. In addition, delta- and epsilonproteobacterial sulfate- and, in some cases, iron(III)-reducers [22] represent the main metabolic bacterial components of the communities.

So far, meta-genomic studies in anoxic sediments and, in particular, in mining-impacted lakes, are rare and only few recent studies have identified abundant key prokaryotes and linked them with essential metabolic processes and environmental adaptations [13]. The objective of our study was to investigate the prokaryotic community inhabiting in the anoxic sediment of the sub-saline shallow lake Laguna de Carrizo, in Central Spain, and highlight the metabolic particularities of this aquatic environment which previously was operated as a gypsum mine. The Carrizo Lake is characterized by an unusual prevalence of Ca^{2+}, Mg^{2+} and SO_4^{2-}, together with a low concentration of other biogenic mono-valent cations (see details in **Methods** section).

Methods

Study site, sampling and DNA extraction

Laguna de Carrizo, located in Madrid (+40° 18' 30.99", -3° 39' 34.70; area approximately 12 km^2; maximum depth 2.4 m; altitude 521 m; **Figure S1**), represents a unique ecosystem in the Central Iberian Peninsula. The area of Carrizo was used since the 17th century to mine gypsum ($CaSO_4 \times 2\ H_2O$) to supply a wide range of industries. In 1977, when the ground water level was reached, the mine was abandoned and the upwelling of subterranean water filled the excavated area. In 1990 the area was declared an abandon industrial site whose restoration is of environmental interest and, since 2004 it belongs to the Drainage and Wetland Regional Catalogue of Madrid (Spain). In Laguna de Carrizo water presents a conductivity of 3160-4910 μS/cm (sub-saline water), a pH of 7.70, a transparency (or light penetration) of 1.8 m and

contains circa 15 g/L of salts. The chemical and mineralogical analyses, done according to Standard Methods (APHA 1998 and ref. [70]), revealed that the sediment contained Ca^{2+} (2.43 – 2.63 g/L), Mg^{2+} (0.40 – 1.49 g/L), Na^{+} (0.09 – 0.19 g/L), K^{+} (0.004-0.04 g/L), NH_4^{+} (989-1249 μg/L) and Fe (0.07-0.11 mg/L). The major ions were SO_4^{2-} (6.96 – 10.89 g/L), $S_2O_3^{2-}$ (3.7-5.0 mg/L), polysulfide (6.5-10.5 μg/L), SO_3^{2-} (2.5-5.7 μg/L), PO_4^{3-} (3.2-3.5 μg/L), Cl^{-} (0.1-0.27 g/L), HCO_3^{-} (0.26-0.42 g/L) and NO_3^{-} (54-744 μg/L). CO_3^{2-}, NO_2^{-}, methane (CH_4; measured by gas chromatography for analysis of gaseous hydrocarbons) and heavy metals (as measured by inductively coupled plasma analysis) were not detectable. Silicate was also found at a concentration ranging from 44 to 100 mg/L. Organic compound analyses indicated that that the sediment contained organosulfonates such as taurine (2-aminoethanesulfonate) (0.14 μg/Kg) and cysteate (2-amino-3-sulfopropionate) (0.68 μg/Kg).

On February 15, 2007, superficial (0 to 20 cm depth) sediment samples (at a depth of 2.4 m) were collected using a Petite Ponar® clamshell-style dredge. The overlaying water was O_2-free, as determined with the Winkler method. The sample was stored at -20ºC until DNA was extracted. DNA was isolated directly from cells previously separated from the environmental matrix. Briefly, suspensions of microbial consortia were obtained by density gradient centrifugation with Nycodenz (Axis-Shield PoC, Norway) as described previously [20]. The resulting cell pellet was subjected to metagenomic DNA extraction using the commercial kit GNOME® DNA (QBIOgene). DNA was visualized by using 0.7% (wt/vol) agarose gel electrophoresis and quantified both spectrophotometrically and with PicoGreen (Molecular Probes, Carlsbad, CA).

Chemicals and enzymes

Chemicals, biochemicals and solvents were purchased from Sigma-Fluka-Aldrich Co. (St. Louis, MO) and were of p.a. (pro analysi) quality. Oligonucleotides for DNA amplification and sequencing were synthesized by Sigma Genosys Ltd. (Pampisford, Cambs, UK). Restriction and modifying enzymes were from New England Biolabs (Beverly, Massachusetts). Ni-NTA His·Bind chromatographic media was from QIAGEN (Hilden, Germany). *E. coli* strains GigaSingles for cloning and BL21(DE3) for expression using the pET-41 Ek/LIC vector (Novagen, Darmstadt, Germany), were cultured and maintained according to the recommendations of the suppliers. All recombinant enzymes used in the present study were PCR-amplified utilizing a PCR-based strategy and custom oligonucleotide primers, cloned, expressed, purified and their kinetic parameters determined as described in **SI Methods.**

Construction of 16S RNA gene clone libraries and clone sequencing

PCR amplification was performed with a serial dilution of DNA template. Bacterial 16S RNA genes were amplified using the bacterial-specific primers F27 (5'-AGAGTTTGATCMTGGCTCAG-3') and R1492 (5'-CGGYTACCTTGTTACGACTT-3'). To analyse in more depth the *Planctomycetes*, we also used Pla f949 (5'-GCGMARAACCTTATCC-3') and Pla r1408 (5'-CCNCNCTTTSGTGGCT-3') that are *Planctomycetes*-specific primers. Archaeal 16S RNA genes were amplified using the and the archaeal-specific primers Ar20F (TTCCGGTTGATCCYGCCRG) and Ar958R (YCCGGGGTTGAMTCCAATT). Amplification was done in a 20 µl reaction volume with recombinant *Taq* DNA Polymerase (Invitrogen, Gemany) and original reagents, according to the basic PCR protocol, with the annealing temperature of 45 and 50°C (bacterial and archaeal rRNA respectively), for 30 cycles. PCR amplicons were purified by electrophoresis in 0.8% (wt/vol) agarose gels, followed by isolation from excised bands using a QIAEX II Gel Extraction Kit (Qiagen, Germany). The purified PCR products were ligated into plasmid vector pGEM (pGEM Cloning kit, Invitrogen, Germany) with subsequent transformation into electrocompetent cells of *E. coli* (TOP 10) (Invitrogen, Germany). Clones of bacterial and archaeal rRNA were sequenced using primers M13 forward (5'-GACGTTGTAAAACGACGGCCAG-3') and M13 reverse (5'-GAGGAAACAGCTATGACCATG-3'), according to the protocol for BigDye Terminator v3.1 Cycle Sequencing Kit from Applied Biosystems (USA). The sequencing reactions were performed using an AB 3730 apparatus from Applied Biosystems (USA).

Phylogenetic analyses of 16S RNA gene sequences

Phylogenetic inference was carried out using the ARB software package [46]. Sequences were automatically aligned using the SINA aligner against SILVA SSURef 100 [60] and LTP s100 [84] reference alignments and manually inspected to correct misplaced bases. To improve resolution at lower taxonomic levels, three independent reference phylogenetic trees were reconstructed, one comprising just members of the phylum *Proteobacteria,* a second with the remaining bacterial phyla, and a third just comprising the domain *Archaea*. The distinct datasets with almost complete SSU sequences were first sieved with a 30% conservational filter, and then the phylogeny was reconstructed with the neighbor-joining algorithm using the Jukes-Cantor correction. The resulting tree topologies were carefully checked against the currently accepted classification of *Prokaryotes* (LPSN, http://www.bacterio.cict.fr) to verify the absence of incongruent phylogenetic relationships.

For this study, the sequences were grouped in OTUs, assuming that one OTU includes sequences with similarity values equal to or higher than 97%, using the software DOTUR [67]. Additionally we considered an OPU [45] to be represented by each single group of clones forming an independent clade in the tree without regarding any rigid similarity cut-off value. Both OTUs and OPUs were plotted to obtain rarefaction curves (**Figure S2**). Statistical analyses were performed using the PAST program.

Library construction and sequencing

Cosmid libraries using the pLAFR3 vector and *E. coli* DH5α were constructed according to Guazzaroni *et al.*, [20]. The cosmid library consisted of 6,500 clones with an average insert size of 29.7 kb (ca. 193 Mbp) that were picked with a QPix2 colony picker (Genetix Co., UK) and grown in 384-microtiter plates containing LB with tetracycline (10.0 μg/ml) and 15% (*v/v*) glycerol and stored at -80°C. Three hundred eight four cosmid clones were randomly selected and fully sequenced with a Roche GS FLX DNA sequencer (454 Life Sciences) (Life Sequencing S.L, Valencia, Spain). Additionally, the full library (6,500 clones) was subjected to functional screens with α-naphtyl acetate (for detecting esterase activity) and o-dianisidine/H_2O_2 (for detecting peroxidase activity) following conditions described elsewhere [81]. Eight clones (two esterase and six peroxidase positives) were selected and further sequenced as a pool with a Roche GS FLX DNA sequencer and the resulting sequence added to that of 384-randomly selected clones.

Assembly was performed by Newbler – tool GS De Novo Assembler v.2.3 (Roche). The estimated error rate: (incorrect bases/total number of expected nucleotides) of 0.49% has been considered for GS20 reads [26]. The error rates for GS20 reads were calculated using the Needleman-Wunsch algorithm [52].

Cosmid sequences analysis

Gene prediction was carried out using the Metagene software [56]. Batch cluster analysis of metagenome sequences was performed with the GenDB v2.2 system [50] by collecting for each predicted ORF observations from similarity searches against sequence databases (nr, SwissProt [3]), KEGG [30], COG [77], genomesDB (see next paragraph) and protein family databases (Pfam [15]) and InterPro [25]). Predicted protein coding sequences were automatically annotated by the software MicHanThi [61]. The MicHanThi software predicts gene functions using a fuzzy logic-based approach based on similarity searches using the NCBI-

nr (including Swiss-Prot) and InterPro database. Further, manual annotation and data mining was performed by using JCoast, version 1.6 [62].

To highlight the phylogenetic consistency, all proteins were searched for similarity by BLAST analysis for the phylogenetic distribution of best hits against genomesDB with a cut off with expectation E value below $1e^{-05}$. Genome DB [62] is a composite database built from the proteome FASTA files obtained from the NCBI Reference Sequences database (RefSeq) for all fully sequenced bacterial and archaeal genomes. Each genome, chromosome, and protein in the file was tagged with a unique internal numerical identifier. In addition, taxonomic and contextual information was parsed from the NCBI Entrez Genome Project database. When available, further contextual data was included pertaining to genome size, guanine-cytosine content, Gram staining, shape, arrangement, endospore formation, motility, salinity, oxygen, habitat and temperature range.

To identify potential metabolic pathways, genes were searched for similarity against the KEGG database. A match was counted if the similarity search resulted in an expectation E value below $1e^{-05}$. All occurring KO (KEGG Orthology) numbers were mapped against KEGG pathway functional hierarchies and statistically analyzed.

MIENS submission

Consistent contextual data acquisition for MIENS compliant submission has been done using the web-based software MetaBar [23].

Results and discussion

General features

On February 15, 2007, 250 g of sediment surface samples were collected from Laguna de Carrizo (details of sampling site and chemical compositions is given in **Methods**). The sample is about twice as salty as seawater, with divalent cations (~2.5 g/L Ca^{2+} and ~0.7 g/L Mg^{2+}) dominating over monovalent (90-190 mg/L Na^{+}, 4-40 mg/L K^{+} and 0.9-1.2 mg/L NH_4^{+}), with 95% being SO_4^{2-}, 3.6% HCO_3^{-} and 1.3-2.3% Cl^{-}; nitrate and phosphate were present at trace concentrations. This ion composition contrasts with that observed in seawater like environments (including common and solar saltern environments) where Na^{+} and Cl^{-} dominated (e.g. [45]), as well as in deep sea environments where Ca^{+2} is found at a ratio 1:200 as compared to the dominant ions [9]. Examples of Ca^{2+} rich environments are the calcium sodium, chloride solution of Soudan Mine [13] and the calcium carbonate ($CaCO_3$) sediments

of profundal lake sediment Lake Kinneret [69]; however, in Carrizo sediment the concentration of SO_4^{2-} is more than 10 times higher than found in those environments and also in subterranean water bodies and aquifers [7]. The Dead Sea represents an example of an environment where Mg^{2+} and Ca^{2+} dominated (albeit their concentration exceed from 7 to 32 times that found in Carrizo lake; [4]) over Na^+ and K^+; however, the sulfate and chloride concentrations in Carrizo lake are 28- and 852-fold lower and higher, respectively, that in the Dead Sea. Therefore, the chemical analysis revealed the unique characteristics of the anoxic sediment of the sub-saline shallow Carrizo lake.

Total DNA was extracted for PCR-based 16S rRNA gene diversity survey of the community structure of the Carrizo sediments. In addition, we generated a pLAFR3 library of about 6,500 clones with an average insert length of 29.7 kbp that was sequenced with a Roche GS FLX DNA sequencer, the approximately total archive of 193 Mb yielded about 92 Mb of raw DNA sequence.

Prokaryotic community structure of the Carrizo lake's sediments

The Good's coverage index [18] for the 106 bacterial OTUs (operational taxonomic unit [86]) and 65 OPUs (operational phylogenetic units [45]) identified in the 195 16S rRNA clones range from 0.65 to 0.86, respectively (**Table 1**). Similar coverage results were obtained for the archaeal clone library (52 sequences) in which the Good's coverage indexes were 0.79 for the 22 OTUs and 0.85 for the 18 OPUs. It is difficult to interpret what the sequence diversity of a given clade means in terms of populations of naturally occurring species [36], but OPUs may be considered from the taxonomic point of view to be equivalent to genera and, in some cases, families from the taxonomic point of view [84]. The indicated level of discrimination proved that the community was diverse (**Figure 1; Table 1**), but also that a satisfactory coverage of the microbial diversity had been achieved in both libraries (see **SI Text** for additional information). The phylogenetic reconstruction showed that the sequences were scattered throughout the whole phylogenetic tree in accordance with the large estimated diversity (**Figures 1**, and **Figures S3**, **S4** and **S5**). It is noteworthy that about 22% of the total rRNAs genes cloned (representing about 20% of the OPUs) showed similarities below 94.5% with any SSU sequences currently deposited in public repositories (either from cultured organisms or from environmental clones).

Among the different bacterial phylotypes recovered, almost two-thirds of the sequences were identified as belonging to the phylum *Proteobacteria* (**Figures 1** and **S3**, and **Table S1**).

The most abundant sequences in the library affiliated with the *Beta*- and *Deltaproteobacteria,* whose sum encompassed nearly 50% of proteobacterial clones (37% of the OPUs). A large fraction of these sequences (42% and 16%, respectively) did not affiliate with any known family, and clustered with branches represented only by uncultured microorganisms, mostly recovered from anaerobic communities in lake and river sediments, in the waters of freshwater reservoirs, wetland soils, as well as calcite, karst and calcite travertine systems [66], microbial mats from aphotic (cave) sulfidic springs [14] and hot springs [38], gold mine water streams [24], acid mine drainage systems, as well as marine sediments [32, 43], and (an)aerobic wastewater digesters [29, 63]. Twenty five percent of all betaproteobacterial OPUs were associated with the highly versatile genus *Burkholderia*, and to less extent to potential chemolithotrophic iron- and sulfur-oxidizing organisms such as *Gallionella* spp. and *Thiobacillus* spp., and putative phototrophs such as *Rhodocyclus* spp. Almost 24% (or 19% of the OPUs) of all clones affiliated with *Deltaproteobacteria*, a class which comprises the major group of sulfate-reducing bacteria (SRB). The most represented SRB sequences affiliated with *Desulfobacteraceae*, *Desulfobulbaceae*, *Syntrophaceae* and *Syntrophobacteraceae*, which together made up circa 75% of the deltaproteobacterial sequences. The third major group of phylotypes detected affiliated with the *Gammaproteobacteria* class (**Figure 1, Table S1**) encompassing 12.8% of the clones (13.8% of the OPUs). Among these sequences a large proportion affiliated with purple sulfur bacteria (*Chromatiaceae*; typical inhabitants of stagnant pools) and versatile heterotrophs such as *Pseudomonas* and *Xanthomonas*-like organisms, followed by sulfur-oxidizing phototrophs such as *Lamprocystis*, and one sequence distantly affiliated to methanotrophic organisms such as *Methylocaldum* (**Figure S3**). They were closely related to communities found in solar salterns [1, 2], waters and sediments of freshwater reservoirs [55, 83], wetland soils as well as karst and phreatic sinkholes and deep-sea marine sediments [82]. Eleven clones (5.6%) affiliated to *Alphaproteobacteria* that were composed in essence of *Sphingomonadaceae* and *Rhodobacteraceae*-like organisms. *Epsilonproteobacteria*, constituting 2.1% of the Carrizo Lake bacterial clones, were affiliated to organisms distantly related to chemolithotrophic *Sulfurovum litotrophicum* and *Sulfurimonas autotrophica*, both involved in the redox sulfur cycle, and to uncultured bacteria from activated wastewater sludges. The remaining sequences in Carrizo Lake bacterial library (most closely related to sequences recovered from freshwater environments, including phreatic sinkholes), which represent about 35% of all the OTUs and OPUs, were related to *Acidobacteria*, *Bacteroidetes*, *Fibrobacteres*, *Firmicutes*, *Lenthisphaerae*, *Nitrospirae*, and to

the candidate divisions JL-ETNP-Z39, OP3, TA06, TM6, WS3 and WS1 with no cultivable organisms (**Table S1**, **Figures 1** and **S4**).

The above phylogenetic analysis of bacterial clone sequences (with *Proteobacteria* being predominant) resulted in overlaps with sequences from other lakes, including saline lakes such as karst and calcite travertine systems, and from (an) aerobic wastewater digesters [5, 29, 63]. The *Proteobacteria* are commonly observed in waters and sediments from other saline and freshwater lakes and thus they do not appear to be specific for Carrizo anoxic sediment. The *Beta*- and *Deltaproteobacteria*, by far the most abundant in Carrizo sediment, appear to be numerically important in anoxic sediment from freshwater lakes, the last one playing a cardinal role in anoxic settings, including anoxic lakes [5, 12, 41]. Based on proximity to cultivated species of known physiology, at least seven different metabolic types could be hypothesized to the *Beta*- and *Deltaproteobacteria* inhabiting the Carrizo anoxic sediment: iron- and sulfur-oxidizing organisms (*Gallionella*- and *Thiobacillus*-like), denitrification bacteria (*Sterolibacterium- and Denitratisoma*-like), sulfate-reducers (*Desulfobacca*-, *Desulfosarcina*- , *Desulfococcus*- and *Desulfocapsa*-like), methylotrophs (*Methyloversatilis*-like), synthrophic bacteria (*Syntrophus*-like, that typically establish interspecies H_2-transfer symbioses with methanogenic archaea [5]) and dehalogenating (*Desulfomonile*-like) and phenol degrading (*Syntrophorhabdus*-like) bacteria. The *Gamma*- and *Alphaproteobacteria* were also abundant, with six clones closely related to cultivated phototrophic sulfur bacteria that oxidize reduced sulfur species (e.g. *Lamprocystis*- and *Thiorhodovibrio*-like) and one to nitrogen-fixing methanotrophs (*Methylocaldum* and *Methylococcus*-like). Therefore, most *Gammaproteobacteria* could in fact be oxidizing H_2S, S^0 or thiosulfate in the Carrizo sediment, as reported also in similar freshwater lakes [5]. Finally, *Epsilonproteobacteria* (most closely related to those found in anaerobic digestors), which are naturally associated with sulfide-rich environments and sulfur spring, were less abundant, thus suggesting that oxidizing sulfide or sulfur capabilities in Carrizo sediments are less represented as compared to other metabolic process. Although, they are absent or rare in common freshwater lakes [11], they appear particularly abundant in oxic/anoxic interfaces (redox clines) in marine environments and suboxic/anoxic lake sediments [5].

By contrast to the previous observations, candidate divisions TA06 and WS1, for which 6 distinct clones were found in our study, appear to be unique in saline lakes and marine sediments. In Carrizo sediment, TA06 formed a cluster with 3 clones related to communities from phreatic sinkholes and three marine (including one estuarine) sediments. WS1-related

clone were closely related to a phylotype retrieved from a phreatic sinkhole and a hypersaline microbial mat [43]. Thus, it appears that sediment conditions, which are considerably distinct from those existing in other freshwater ecosystems, could explain the presence of TA06 and WS1 members in the anoxic sediment herein investigated. Unfortunately, since there are no cultured representatives related to our clones, their physiology (e.g. in relation to salinity) remains unknown.

Most of the 52 sequenced archaeal clones (88.5%) affiliated with *Euryarchaeota*, and encompassed 15 OPUs (**Table S1** and **Figures 1** and **S5**). Only three OPUs (6 distinct clones) affiliated with the uncultured *Crenarchaeota* groups Marine Benthic Group B (MBG-B) and Miscellaneous Crenarchaeotic Group (MCG), thus appearing that this archaeal clade plays a minor role in this habitat. The sequences of the first cluster were closely related to uncultured archaeon clones recovered from an anaerobic sludge digestor [63] and a low-pH (≤4) minerotrophic fen [6]. The sequences in the second cluster were related to uncultured Crenarchaeota from hypersaline microbial mat [64] and sulfuric rich submerged sinkhole ecosystems. Among the *Euryarchaeota* sequences, only three OPUs represented by five clones could be affiliated with potential methanogenic *Archaea*. From these, only two clones were related to *Methanobacteria* typically detected in the anoxic sediments at the bottom of ponds and marshes [87] whereas one clone was most closely related to the methanogenic genus *Methanosaeta*, frequently detected both in anaerobic methane-producing bioreactors and in shallow marine sediments rich in methane [47, 79]. The low number of methanogenic *Archaea* identified, together with the fact that no clone sequences recovered matched closely to known sequences recovered from methanogenic sediments further indicated that methanogenesis might be a minor metabolic process in Carrizo lake, as compared to common anoxic lake environments (e.g. [12, 69, 85]), where this process dominated. This agrees with previous observations in saline and alkaline soda lakes where the high sulfate and salt concentrations repressed autotrophic methanogens while promoting active sulfur cycle (e.g. [72]).

However, the largest set of *Euryarchaeota* sequences (comprising 41 sequences and 12 OPUs) affiliated with the uncultured *Thermoplasmatales* CCA47 group, for which two clusters were identified. Twenty seven and 14 sequences formed the first and second cluster, and they were related to archaeal communities of a variety of marine sediments [33], iron- and sulfur-precipitating microbial mats at submarine mud Volcano [57] and microbial mats of hypersaline coastal lagoons [27], deep sinkhole ecosystems and salt marine marsh sediments [53, 54),

respectively. Thus, it appears that *Thermoplasmatales* CCA47 sequences belong to organisms highly adapted to conditions existing in saline, but not common freshwater, ecosystems.

Taken together, whereas many bacterial clones in our study were most closely related to sequences recovered from other freshwater and marine environments and, to minor extent, to sequences from anaerobic wastewater digesters (**Figures S3** and **S4**), the composition of archaeal clones showed remarkable differences. Thus, to the best of our knowledge, the presence of *Thermoplasmatales* CCA47 group in freshwater ecosystems (including anoxic sediments) has not been reported. Although, Schwarz *et al*. [69]) and Glissmann *et al*. [17] reported the presence of *Thermoplasmales* relatives of the Marine Archaea Group III (but not CCA47 group) in anoxic sediments of subtropical and eutrophic profundal lakes, those constitute a minor component of the archaeal community (below circa 17%), which was dominated by common *Methanomicrobiales* and *Methanomicrobiaceae*. The relatively close relationship of the CCA47 group with a group of cultured acidophilic and cell wall-less *Archaea*, also belonging to *Thermoplasmatales*, contrasts with the neutral and slightly alkaline nature of the pore waters of the lake. Unfortunately, due to the lack of cultivable members within the Euryarchaeota clades detected in the Carrizo Lake, little is known about the mechanisms by which CCA47 and also, MBG-B and MCG, like organisms obtain energy, although, they are exclusively found in saline and oxygen-depleted locations [5, 28, 65, 78].

Taxonomic classification of the metagenome sequences (see **Methods** section and **Table S2**) was mostly in line with 16S tag analysis for known taxa (**Figure S6**). However, it should be noticed that this pipeline cannot identify poorly studied taxa without known references sequences or protein-coding genes as they occur in particular in the Carrizo sediment (e.g. candidate divisions TA06 and WS1 and *Thermoplasmatales* CCA47 group). Despite this limitation, the taxonomic binning of the metagenome confirmed the rRNA-based observations with the dominance of proteobacterial related sequences (see **SI Text** for additional information).

Functional signatures for sulfur and nitrogen metabolisms

Freshwater ecosystems are a general focus of intense research, but most of this interest is centred on their phylogeny (using 16S rRNA sequence analysis and related techniques), contrasting with the limited information about the gene inventory via meta-genomic [10, 13], which may shed light on microbial ecology of distinct ecosystems. Here, the metabolic

potential that is expected to be present in Carrizo lake sediments was analysed, with a particular focus in the sulfur and nitrogen associated processes (see details in **Tables S3-S5**).

The metagenome contained 1333 assembled contigs (71 of them with lengths from 10 to 30 kbp) representing 3,690 CDS (coding sequences) with an average read length of 633 bp (**Table S2**, **Figure S7**). The G+C content of each CDS was calculated, and the values were normally distributed between 78.8% and 14.7%, with a mean of 53.9% for the library. With a maximum E value criterion of 10^{-5}, 22% of the sequences in this metagenome library did not have any sequence similarity (hypothetical proteins) and another 19% (686 sequences) were similar to proteins of unknown function (conserved hypotheticals). Thus, an important fraction of this ecosystem remains unknown and its metabolism is difficult to be unraveled.

A total number of 46 (or 2.1% of hits with assigned function) genes coding enzymes potentially involved in the sulfur cycle were identified. As shown in **Figure 3** it is apparent that Carrizo lake community likely utilizes the dissimilatory phosphoadenosine phosphosulfate (PAPS) reductase system to convert sulfate (SO_4^{2-}) to sulfite (SO_3^{2-}), the NrfD polysulfide- and MopB thiosulfate-reductase-like systems to produce sulfide (S^{2-}) from polysulfide (S_n^{2-}), thiosulfate ($S_2O_3^{2-}$) and thiosulfonate, and the sulfide oxidoreductase (SQR-like) to oxidize S^{2-} to S_n^{2-} that are again substrates for polysulfide reductases. Experimental proofs are provided in **SI Text** and **Tables S3-S5**. No evidence for the assimilatory sulfate reduction (by APS reductases) nor direct conversion of SO_3^{2-} to S^{2-} by dissimilatory sulfite reductases (Dsr) and (thio-) sulfate-oxidation (mediated by Sox multi-enzyme complex) was found; however, the possibility of a metagenome biass (low genome coverage) cannot be ruled out, because the presence of genes coding the first two enzymes has been demonstrated by PCR-based approaches using degenerated primers [37]. The identification of five thiosulfate:cyanide sulfurtransferases (RhoD), two aryl sulfotransferases (see **SI Text** and **Table S4** for experimental evidences) and 2 sulfatases (EC 3.1.6.-) is also supportive for the active utilization of thiosulfates and (aryl) sulfate esters as sources of sulfite and sulfate, respectively. Further, a YedY-like sulfite oxidase, involved in the reduction of linear and cyclic sulfoxides, organosulfonates and N-oxides [31], but lacking sulfite-oxidizing activity, was identified and confirmed experimentally (see **SI Text** and **Table S4**), which suggests that these types of compounds (detected in the sediments by chemical analysis) can potentially be used as sulfur source by Carrizo community members. To the best of our knowledge, this is the first time to report YedY-like sulfite oxidase activity in an anoxic (including both freshwater and marine) environment.

Forty one genes (or circa 2.0% of hits with assigned function) coding enzymes potentially involved in the assimilation and transformation of N-sources, namely diatomic nitrogen (N_2) (nitrogen fixation), nitrate (NO_3^-), nitrite (NO_2^-), ammonium (NH_4^+), as well as N-oxides and nitro-, nitrile- and cyanide-substituted (including aromatic) compounds (widely distributed in environments associated with industrial wastewater and residual agricultural chemicals [34]), were identified (**Figure 4** and **Tables S3-S5**). They include 5 NifX/B-like dinitrogenases, 6 IscU/NifU-proteins related to dinitrogen fixation, 2 nitrate/nitrite transporter (NarK), 6 nitrate reductase-like proteins (NarG,H, I, J,) (see **SI Text** and **Table S4** for experimental evidences), 1 nitrite reductase (NrfC), 4 nitro-reductases (3 of them experimentally characterized; **SI Text** and **Table S4**), 2 nitrilases/cyanide hydratases (1 experimentally characterized; **SI Text** and **Table S4**) and three 2-nitropropane-like dioxygenases potentially involved in nitrite production from nitropropane (1 experimentally characterized; **SI Text** and **Table S4**). Additionally, a QueF like nitrile reductase likely responsible of the direct NADPH-dependent biological reduction of nitrile functional groups to a primary amine (rarely been observed in biological systems [40]), was identified and further characterized experimentally (**SI Text** and **Table S4**). The presence of nitropropane dioxygenase activity was unexpected for an anaerobic environment and might be explained by either sedimentation of genomic debris from the interface, or by the presence of symbiotic bacteria-eukaryote associations as reported previously for a RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase) protein in a deep-sea ecosystem [49]. Finally, we detected a number of nitrogen regulatory proteins plus a number of phosphotransferase systems (9 hits in total, coverage >30%), which possibly play a role in nitrogen assimilation [58].

It cannot be excluded that other enzymes relevant for the sulfur and nitrogen cycle slipped detection due to low meta-genome coverage; however, the gene inventory herein provided complement the metabolic activities suggested by the 16S rRNA (based on proximity to cultivated species of known physiology). In fact, the meta-genomic data suggested the operation of a sulfur cycle ($HS^- \rightarrow S_n^{2-} \rightarrow HS^-$) and, moreover, that thiosulfate and thiosulfonate-reducing bacteria are positioned at a decisive stage. Finally, the identification of rhodanases and nitrilases (whose presence cannot be suggested by 16S rRNA tags), suggested that functions for these enzymes may not be only cyanide detoxification (whose presence we were not able to detect *in situ*), but that they may be related to the production of ammonium (NH_4^+) and SO_3^{2-}, thus contributing to the S- and N- cycles, in contrast to other saline ecosystems [37, 72].

We further investigate the presence of sulfur/nitrate assimilation clusters as they may be selectively favorable because it facilitates the coordinated expression of the constituent genes [51, 71]. We further perform tentative taxonomic assignments based on BLAST hits. As a result of this analysis, four assembled contigs were found to possess genes coding proteins for the assimilation of both S- and N-sources. Briefly, the 8,446-bp long contig cLDC0361 (45.09% GC content) contains a full set of genes which encode proteins required for the biosynthesis of sulfur-containing aminoacids, cysteine and methionine, linked with a nitrogen-fixing NifU domain protein (LDC_0888). The majority of the genes in cLDC0361 were most similar to *Syntrophus* members, the sixth most represented microorganism in the Carrizo community. Additionally, the 11,467-bp long cLDC0380, which possesses a much lower GC content (34.23%), appears to encode for enzymes of the sulfur cycle, namely, the reduction of (thio) sulfate to hydrogen sulfide by LDC_1013 and polysulfide to sulfide by LDC_1015 (for which experimental evidences are given in **SI Text** and **Table S4**), as well as enzymes for the assembly and activation of the NifU nitrogenase catalytic components (i.e. the [Fe-S] cluster and the molybdopterin co-factor). cLDC0380 was found to be highly syntenic to *Epsilonproteobacteria* (whose sequences encompass 2.9% of the total 16S riboclones (**Table S1**) and 9% of the total BLAST hits (**Figure S6**) of the metagenome) with 41% and 25% of all genes belonging to *Wolinella succinogenes* DSM 1740 and *Arcobacter butzleri* RM 4018, respectively. The 16,889-bp long contig cLDC0376 (32.46% GC content) appears to encode for two NifBX dinitrogenase iron-molybdenum cofactor biosynthesis proteins (LDC_0986 and LDC_0987), two cobyrinic acid a,c-diamide synthases that use glutamine or ammonia as a nitrogen source for the anaerobic biosynthesis of vitamin B12 [16] and a polysulfide-sulfur transferase (LDC_0978). The DNA fragment showed similar genomic organization as their counterparts from N_2-fixing and H_2S-oxidizing *Sulfurovum* sp. NBC37-1 (28% or 5 hits) and *Sulfurospirillum deleyianum* DSM 6946 (22% or 4 hits). The gene organization in this contig in relation to genomic fragments from both chemolitotrophic sulfur-respiring *Epsilonproteobacteria* is highlighted in **Figure 5**. Finally, the 20-Kbp long cLDC0001 has two differentiated gene clusters characterized by their atypical GC content (35.68% *versus* 58.04%) and the presence of numerous genes with high similarity to genes found in distantly related species (**Figure 5**). The high GC containing island (position 11,600-20,012) bears a block of clustered genes encoding two NarK-like high affinity nitrate/nitrite transporters (LDC_0007 and LDC_0008), the alpha and beta subunits of a respiratory nitrate reductase which catalyses the reduction of nitrate to nitrite (LDC_0009 and LDC_0010, providing experimental evidences), and a chaperone required for the proper folding

of the nitrate reductase (LDC_0011). Most proteins from this high GC island were most closely related to those of *Albidiferax ferrireducens* (formerly *Rhodoferax ferrireducens*), an anaerobic proteobacterium (beta subdivision) with Fe(III) reducing capabilities, thus suggesting the presence of such metabolism in an *Albidiferax*-like bacterium inhabiting Carrizo sediment. In this context, the dissimilatory reduction of iron has been shown to be an important biochemical process in anoxic, mining-impacted lake sediments [8]. Upstream of this block, the genomic fragment at position 1,150-10,449 has a GC content of 35.68% and encodes a number of hypothetical proteins with no clear taxonomic affiliation.

The above data suggest that horizontal gene exchange between different members of the bacterial community and phage integration (e.g. cLDC0376 contains three transposases and one phage integrase) may be highly active in the Carrizo community and, moreover, they may play important roles in the sulfur and nitrogen cycling. This may agree with the observation that in marine sub-saline systems horizontal gene exchange between different members of the prokaryotic communities is highly active, thus favoring adaptive evolution [36, 75]. Moreover, the above analysis demonstrated that representatives of synthrophic bacteria (e.g. *Syntrophus*-like) and *Epsilonproteobacteria* are major contributors of the sulfur and nitrogen cycling in Carrizo sediments.

Conclusions

In this study, cultivation-independent metagenomic and 16S rRNA assessments were used to infer correlations between systems performance and phylogenetic and relevant genomic capacities in the microbial community inhabiting the anoxic sediment of a sub-saline shallow lake (Laguna de Carrizo), initially operated as a gypsum mine. Compared to other saline and freshwater ecosystems described to date Carrizo Lake is characterized by an unusual ionic composition. The information retrieved agrees with the expected assemblage of organisms thriving in anoxic sediments; our study gives a comprehensive insight into the structure of the bacterial and archaeal community of a shallow anoxic lake, indicating that thiosulfate- and thiosulfonate-reducers, sulfate-reducers and iron-oxidizing, sulfur-oxidizing, denitrification, synthrophic and phototrophic sulfur bacteria are of particular importance in Carrizo sediment as compared to methanogens (predominant in common anoxic freshwater sediments). Genome data herein provided suggests that (thio) sulfates and (thio) sulfonates, polysulfides, sulfoxides and organosulfonates, together with nitro-, nitrile- and cyanide-substituted compounds might be major primary sources of biological sulfur and nitrogen in this niche.

These metabolic capacities have rarely been observed together in open marine, sub-saline or freshwater environments. It is likely that microorganisms in Laguna de Carrizo sediments experience episodes of extreme sulfur/nitrogen-like (including toxic) stress/pressure, where transfer of complete assimilation pathways (possibly to improve microbial fitness) is an active mechanism. Results suggest that the anthropogenic activities around the Carrizo area may have exerted strong selective pressure on the microbial community to adapt it to toxic chemicals (major abiotic stressors). Since most of the BLAST hits were associated with the *Sulfurovum* genus, the results suggest that members related to this genus might be highly active within the Carrizo community, thus opening new research opportunities to further investigate their metabolic arsenal. This should be of interest due to the limited genomic information described to date in anoxic saline environments [13]. Furthermore, to our knowledge this is the first report of *Thermoplasmatales* CCA47 group in anoxic shallow sediments and, freshwater ecosystems, in general, and our data indicate that these members constitute a prevalent component of the Carrizo archaeal community, as compared to what was previously described in similar habitats. The fact that members of this group (together with bacterial candidate divisions TA06 and WS1) have been only found in marine and oxygen free environments, suggest salinity as a major determinant for their presence and/or abundance in Carrizo sediment. Further investigations will be required to ascertain their global metabolic role in the overall community and sediment characteristics. It should be noticed that, in addition to salinity, other environmental differences (e.g. carbon supply, sediment redox conditions, sediment depth and relative proportion of ions) may help to explain the observed archaeal diversity patterns in Carrizo lake.

Acknowledgments

This research was supported by the Spanish CSD2007-00005 project and FEDER funds. M-E.G. thanks the CSIC for a JAE fellowship. We also thank Prof. Michail M. Yakimov and Dr. Terence J McGenity for assistance in diversity and adaptation analyses.

References

1. Baati H, Guermazi S, Amdouni R, Gharsallah N, Sghir A, Ammar E (2008) Prokaryotic diversity of a Tunisian multipond solar saltern. Extremophiles 12:505-518
2. Baati H, Guermazi S, Gharsallah N, Sghir A, Ammar E (2010) Novel prokaryotic diversity in sediments of Tunisian multipond solar saltern. Res Microbiol 161:573-582

3. Bairoch A, Boeckmann B, Ferro S, Gasteiger E (2004) Swiss-Prot: Juggling between evolution and stability. Brief Bioinform 5:39-55
4. Bodaker I, Sharon I, Suzuki MT, Feingersch R, Shmoish M, Andreishcheva E, Sogin ML, Rosenberg M, Maguire ME, Belkin S, Oren A, Béjà O (2010) Comparative community genomics in the Dead Sea: an increasingly extreme environment. ISME J 4:399-407
5. Briée C, Moreira D, López-García P (2007) Archaeal and bacterial community composition of sediment and plankton from a suboxic freshwater pond. Res Microbiol 158:213-227.
6. Cadillo-Quiroz H, Yashiro E, Yavitt JB, Zinder SH (2008) Characterization of the archaeal community in a minerotrophic fen and terminal restriction fragment length polymorphism-directed isolation of a novel hydrogenotrophic methanogen. Appl Environ Microbiol 74:2059-2068.
7. Coupland K, Johnson DB (2004) Geochemistry and microbiology of an impounded subterranean acidic water body at Mynydd Parys, Anglesey, Wales. Geobiology 2:77-86
8. Cummings DE, March AW, Bostick B, Spring S, Caccavo F Jr, Fendorf S, Rosenzweig RF (2000) Evidence for microbial Fe(III) reduction in anoxic, mining-impacted lake sediments (Lake Coeur d'Alene, Idaho). Appl Environ Microbiol 66:154-162
9. De Lange GJ, Middelburg JJ, Van der Weijen CH, Catalano G, Luther GW, Hydes DJ, Woittiez JRW, Klinkhammer GP (1990) Composition of anoxic hypersaline brines in the Tyro and Bannock basins, eastern Mediterranean. Marine Chemistry 31:63-88
10. Debroas D, Humbert J-F, Enault F, Bronner G, Faubladier M, Cornillot E (2009) Metagenomic approach studying the taxonomic and functional diversity of the bacterial community in a mesotrophic lake (Lac du Bourget – France). Environ Microbiol 11:2412-2424
11. Donachie SP, Hou S, Lee KS, Riley CW, Pikina A, Belisle C, Kempe S, Gregory TS, Bossuyt A, Boerema J, Liu J, Freitas TA, Malahoff A, Alam M (2004) The Hawaiian Archipelago: a microbial diversity hotspot. Microb Ecol 48:509-520
12. Dong H, Zhang G, Jiang H, Yu B, Chapman LR, Lucas CR, Fields MW (2006) Microbial diversity in sediments of saline Qianghai lake, China: linking geochemical controls to microbial ecology. Microbial Ecology 51:65-82.
13. Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, Saar MO, Alexander S, Alexander EC Jr, Rohwer F (2006) Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics 7:57
14. Engel AS, Porter ML, Stern LA, Quinlan S, Bennett PC (2004) Bacterial diversity and ecosystem function of filamentous microbial mats from aphotic (cave) sulfidic springs dominated by chemolithoautotrophic "*Epsilonproteobacteria*". FEMS Microbiol Ecol 51:31-53
15. Finn RD, Tate J, Mistry J, Coggill PC, Sammut JS, Hotz HR, *et al*. (2008) The Pfam protein families database. *Nucleic Acids Res Database Issue* 36:D281-D288

16. Fresquet V, Williams L, Raushel FM (2004) Mechanism of cobyrinic acid a,c-diamide synthetase from *Salmonella typhimurium* LT2. Biochemistry 43:10619-10627.
17. Glissmann K, Chin K.-J., Casper P, Conrad R (2004) Methanogenic pathway and archaeal community structure in the sediment of eutrophic Lake Dagow: effect of temperature. Microbial Ecology 48:389-399.
18. Good IJ (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40:237-264
19. Gough HL, Stahl DA (2011) Microbial community structures in anoxic freshwater lake sediment along a metal contamination gradient. ISME J 5:543-558.
20. Guazzaroni M.E, Golyshin PN, Ferrer M (2010) Molecular methods to study complex microbial communities. *In: Metagenomics:* Theory, Methods and Applications (Ed. Diana Marco), Caister Academic Press, pp. 55-77
21. Haaijer SC, Harhangi HR, Meijerink BB, Strous M, Pol A, Smolders AJ, Verwegen K, Jetten MS, Op den Camp HJ (2008) Bacteria associated with iron seeps in a sulfur-rich, neutral pH, freshwater ecosystem. ISME J 2:1231-1242
22. Handley KM, Boothman C, Mills RA, Pancost RD, Lloyd JR (2010) Functional diversity of bacteria in a ferruginous hydrothermal sediment. ISME J 4:1193-1205.
23. Hankeln W, Buttigieg PL, Fink D, Kottmann R, Yilmaz P, Glöckner FO (2010) MetaBar - a tool for consistent contextual data acquisition and standards compliant submission. BMC Bioinformatics 11:358
24. Hirayama H, Takai K, Inagaki F, Yamato Y, Suzuki M, Nealson KH, Horikoshi K (2005) Bacterial community shift along a subsurface geothermal water stream in a Japanese gold mine. Extremophiles 9:169-184
25. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, *et al* (2009) InterPro: the integrative protein signature database. Nucleic Acids Res Database Issue 37:D224-228.
26. Huse SM, Huber JA, Morrison HG, Sogin ML, Weich DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 8:R143
27. Jahnke LL, Orphan VJ, Embaye T, Turk KA, Kubo MD, Summons RE, DES Marais DJ (2008) Lipid biomarker and phylogenetic analyses to reveal archaeal biodiversity and distribution in hypersaline microbial mat and underlying sediment. Geobiology 6:394-410
28. Jiang H, Dong H, Zhang G, Yu B, Chapman LR, Fields MW (2006) Microbial diversity in water and sediment of Lake Chaka, an athalassohaline lake in northwestern China. Appl Environ Microbiol 72:3832-3845
29. Juretschko S, Loy A, Lehner A, Wagner M (2002) The microbial community composition of a nitrifying-denitrifying activated sludge from an industrial sewage treatment plant analyzed by the full-cycle rRNA approach. Syst Appl Microbiol 25:84-99.

30. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, *et al* (2008) KEGG for linking genomes to life and the environment. Nucleic Acids Res 36:D480-D484
31. Kappler U (2008) Bacterial sulfite-oxidazing enzymes – enzymes for chemolithotrophs only? C. Dahl and C.G. Friedrich (eds.), Microbial Sulfur Metabolism. Springer 2008, pp. 151.
32. Kendall MM, Liu Y, Boone DR (2006) Butyrate- and propionate-degrading syntrophs from permanently cold marine sediments in Skan Bay, Alaska, and description of *Algorimarina butyrica* gen. nov., sp. nov. FEMS Microbiol Lett 262:107-114
33. Kendall MM, Wardlaw GD, Tang CF, Bonin AS, Liu Y, Valentine DL (2007) Diversity of Archaea in marine sediments from Skan Bay, Alaska, including cultivated methanogens, and description of *Methanogenium boonei* sp. nov. Appl Environ Microbiol 73:407-414
34. Kobayashi M, Shimizu S (2000) Nitrile hydrolases. Curr Opin Chem Biol 4:95-102.
35. Konstantinidis K, DeLong E.F (2008) Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. ISME J 2:1052-1065
36. Konstantinidis KT, Braff J, Karl DM, DeLong EF (2009) Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. Appl Environ Microbiol 75:5345-5355
37. La Cono V, Smedile F, Bortoluzzi G, Arcadi E, Maimone G, Borghini M, Oliveri E, Mazzola S, L'Haridon S, Genovese L, Ferrer M, Giuliano L, Golyshin PN, Yakimov V (2011) Unveiling microbial life of new deep-sea hypersaline lake Thetis. Part 1: Prokaryotes and environmental settings. Environ Microbiol, In press.
38. Lau MC, Aitchison JC, Pointing SB (2009) Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. Extremophiles 13:139-149
39. Lauro FM, Demaere MZ, Yau S, Brown MV, Ng C, Wilkins D, Raftery MJ, Gibson JA, Andrews-Pfannkoch C, Lewis M, Hoffman JM, Thomas T, Cavicchioli R (2011) An integrative study of a meromictic lake ecosystem in Antarctica. ISME J 5:879-895.
40. Lee BW, Van Lanen SG, Iwata-Reuyl D (2007) Mechanistic studies of *Bacillus subtilis* QueF, the nitrile oxidoreductase involved in queuosine biosynthesis. Biochemistry 46:12844-12854
41. Lehours AC, Bardot C, Thenot A, Debroas D, Fonty G (2005) Anaerobic microbial communities in Lake Pavin, a unique meromictic lake in France. Appl Environ Microbiol 71:7389-7400
42. Lehours AC, Evans P, Bardot C, Joblin K, Gérard F (2007) Phylogenetic diversity of archaea and bacteria in the anoxic zone of a meromictic lake (Lake Pavin, France). Appl Environ Microbiol 73:2016-2019
43. Ley RE, Harris JK, Wilcox J, Spear JR, Miller SR, Bebout BM, Maresca JA, Bryant DA, Sogin ML, Pace NR (2006) Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. Appl Environ Microbiol 72:3685-3695
44. Llirós M, Casamayor EO, Borrego C (2008) High archaeal richness in the water column of a freshwater sulfurous karstic lake along an interannual study. FEMS Microbiol Ecol 66:331-342

45. López-López A, Yarza P, Richter M, Suárez-Suárez A, Antón J, Niemann H, Rosselló-Móra R (2011) Extremely halophilic microbial communities in anaerobic sediments from a solar saltern. Environ Microbiol (in press)
46. Ludwig W Strunk O, Westram R, Richter L, Meier, H, Yadhukumar BA, *et al* (2004) ARB: a software environment for sequence data. Nucleic Acids Res 32:1363-1371
47. MacLeod FA, Guiot SR, Costerton JW (1990). Layered structure of bacterial aggregates produced in an upflow anaerobic sludge bed and filter reactor. Appl Environ Microbiol 56:1598-1607
48. Mallet C, Basset M, Fonty G, Desvilettes C, Bourdier G, Debroas D (2004) Microbial population dynamics in the sediments of a eutrophic lake (Aydat, France) and characterization of some heterotrophic bacterial isolates. Microb Ecol 48:66-77
49. Markert S, Arnd C, Felback H, Becher D, Sievert SM, et al (2007) Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. Science 315:247-250
50. Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, et al (2003): GenDB - an open source genome annotation system for prokaryote genomes. Nucleic Acids Res 31:2187-95
51. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:386
52. Needleman SB, Wunsch CD: *J Mol Biol* 1970, 48:443-453
53. Nelson DM, Ohene-Adjei S, Hu FS, Cann IK, Mackie RI (2007) Bacterial diversity and distribution in the holocene sediments of a northern temperate lake. Microb Ecol 54:252-263
54. Nelson KA, Moin NS, Bernhard AE (2009) Archaeal diversity and the prevalence of Crenarchaeota in salt marsh sediments. Appl Environ Microbiol 75:4211-4215
55. Nercessian O, Noyes E, Kalyuzhnaya MG, Lidstrom ME, Chistoserdova L (2005) Bacterial populations active in metabolism of C1 compounds in the sediment of Lake Washington, a freshwater lake. Appl Environ Microbiol 71:6885-6899
56. Noguchi H, Park J, Takagi T (2006). MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. Nucleic Acids Res 34:D5623-30
57. Omoregie EO, Mastalerz V, de Lange G, Straub KL, Kappler A, Røy H, Stadnitskaia A, Foucher JP, Boetius A (2008) Biogeochemistry and community composition of iron- and sulfur-precipitating microbial mats at the Chefren mud volcano (Nile Deep Sea Fan, Eastern Mediterranean). Appl Environ Microbiol 74:3198-3215
58. Pflüger K, de Lorenzo V (2007) Growth-dependent phosphorylation of the PtsN (EIINtr) protein of *Pseudomonas putida*. J Biol Chem 282:18206-18211

59. Porat I, Vishnivetskaya TA, Mosher JJ, Brandt CC, Yang ZK, Brooks SC, Liang L, Drake MM, Podar M, Brown SD, Palumbo AV (2010) Characterization of archaeal community in contaminated and uncontaminated surface stream sediments. Microb Ecol 60:784-795
60. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188-7196
61. Quast C (2006). MicHanThi - Design and Implementation of a System for the Prediction of Gene Functions in Genome Annotation Projects. Diploma work, Universität Bremen.
62. Richter M, Lombardot T, Kostadinov I, Kottmann R, Peplies J, Duhaime MB, Glöckner FO (2008) JCoast – A biologist-centric software tool for data mining and comparison of prokaryotic (meta)genomes. BMC Bioinformatics 9:177
63. Rivière D, Desvignes V, Pelletier E, Chaussonnerie S, Guermazi S, Weissenbach J, Li T, Camacho P, Sghir A (2009) Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. ISME J 3:700-714
64. Robertson CE, Spear JR, Harris JK, Pace NR (2009) Diversity and stratification of archaea in a hypersaline microbial mat. Appl Environ Microbiol 75:1801-1810
65. Röling WF, Couto de Brito IR, Swanell RPJ, Head IM (2004) Response of archaeal communities in beach sediments to spilled oil and bioremediation. Appl Environ Microbiol 70:2614-2620
66. Sahl JW, Schmidt R, Swanner ED, Mandernack KW, Templeton AS, Kieft TL, Smith RL, Sanford WE, Callaghan RL, Mitton JB, Spear JR (2008) Subsurface microbial diversity in deep-granitic-fracture water in Colorado. Appl Environ Microbiol 74:143-152.
67. Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 71:1501-1506
68. Schwarz JI, Lueders T, Eckert W, Conrad R (2007a) Identification of acetate-utilizing Bacteria and Archaea in methanogenic profundal sediments of Lake Kinneret (Israel) by stable isotope probing of rRNA. Environ Microbiol 9:223-237
69. Schwarz JIK, Eckert W, Conrad R (2007b) Community structure of *Archaea* and *Bacteria* in a profundal lake sediment Lake Kinneret (Israel). Sys Appl Microbiol 30:239-254.
70. Siegel De Ea (1999) Microscopía óptica y caracterización de minerales. Departamento Ingeniería en Minas, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile, pp. 98
71. Slot JC, Hibbett DS (2007) Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study. PLoS One 10:e1097
72. Sorokin DY, Rusanov II, Pimenov NV, Tourova TP, Abbas B, Muyzer G (2010) Sulfidogenesis under extremely haloalkaline conditions in soda lakes of Kulunda Steppe (Altai, Russia). FEMS Microbiol Ecol 73:278-290

73. Swan BK, Ehrhardt CJ, Reifel KM, Moreno LI, Valentine DL (2010) Archaeal and bacterial communities respond differently to environmental gradients in anoxic sediments of a California hypersaline lake, the Salton Sea. Appl Environ Microbiol 76:757-768
74. Tamaki H, Sekiguchi Y, Hanada S, Nakamura K, Nomura N, Matsumura M, Kamagata Y (2005) Comparative analysis of bacterial diversity in freshwater sediment of a shallow eutrophic lake by molecular and improved cultivation-based techniques Appl Environ Microbiol 71:2162-2169
75. Tamames J, Mira A (2010) Horizontal gene transfer in prokaryotic microbial communities. In: Metagenomics: Theory, Methods and Applications (Ed. Diana Marco), Caister Academic Press, pp. 39-53
76. Tang X, Gao G, Qin B, Zhu L, Chao J, Wang J, Yang G (2009) Characterization of bacterial communities associated with organic aggregates in a large, shallow, eutrophic freshwater lake (Lake Taihu, China). Microb Ecol 58:307-322.
77. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, et al (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 29:22-28
78. Teske A, Sorensen KB (2008) Uncultured archaea in deep marine subsurface sediments: have we caught them all? ISME J 2:3-18
79. Thomsen TR, Finster K, Ramsing NB (2001). Biogeochemical and molecular signatures of anaerobic methane oxidation in a marine sediment. Appl Environ Microbiol 67:1646-1656
80. Tsiamis G, Katsaveli K, Ntougias S, Kyrpides N, Andersen G, Piceno Y, Bourtzis K (2008) Prokaryotic community profiles at different operational stages of a Greek solar saltern Res Microbiol 159:609-627
81. Vieites JM, Guazzaroni ME, Beloqui A, Golyshin PN, Ferrer M (2010) Molecular methods to study complex microbial communities. Methods Molecular Biology 668:1-37
82. Wang J, Vollrath S, Behrends T, Bodelier PL, Muyzer G, Meima-Franke M, Den Oudsten F, Van Cappellen P, Laanbroek HJ (2011) Distribution and diversity of Gallionella-like neutrophilic iron oxidizers in a tidal freshwater marsh. Appl Environ Microbiol 77:2337-2344.
83. Wu X, Xi W, Ye W, Yang H (2007) Bacterial community composition of a shallow hypertrophic freshwater lake in China, revealed by 16S rRNA gene sequences. FEMS Microbiol Ecol 61:85-96.
84. Yarza P, Richter M, Peplies J, Euzéby J, Amann R, Schleifer K-H, et al (2008) The All-Species Living Tree Project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. Syst Appl Microbiol 31:241-250
85. Ye W, Liu X, Lin S, Tan J, Pan J, Li D, Yang H (2009) The vertical distribution of bacterial and archaeal communities in the water and sediment of Lake Taihu. FEMS Microbiol Ecol 70:107-120.

86. Zaballos M, López-López A, Øvreås L, Galán-Bartual S, D'Auria G, Alba-Casado J, et al (2006) Comparison of prokaryotic diversity at offshore oceanic locations reveals a different microbiota in the Mediterranean Sea. FEMS Microbiol Ecol 56:389-405
87. Zellner G Stackebrandt E, Messner P, Tindall BJ, Conway de Macario E, Kneifel H, Sleytr UB, Winter J (1989) *Methanocorpusculaceae* fam. nov., represented by *Methanocorpusculum parvum, Methanocorpusculum sinense* spec. nov. and *Methanocorpusculum bavaricum* spec. nov. Arch Microbiol 151:381-390

Table 1 Statistical indexes

	Bacteria		Archaea	
	OTUs*	**OPUs**	**OTUs***	**OPUs**
Number of sequences	195		52	
Good's coverage value	0.65	0.86	0.79	0.85
Number of taxa	106	65	22	18
Shannon-Weiner index	4.47	3.78	2.66	2.22
Equitability	0.96	0.91	0.87	0.77

* Clustered at 97% identity

The PAST software v1.82b was used to compute the statistical indexes for the archaeal and bacterial sequences. The formulas used are as follows: Shannon-Weiner index: $H = -\Sigma(ni/nt)Ln(ni/nt)$, where *ni* is the number of sequences of a particular OTU and *nt* is the total number of sequences. Equitability: $J = H'/H'max$, where *H'* is the observed diversity and $H'max = LnS$ or the maximum possible diversity for a sample of *S* equally abundant species. Good's coverage: $C = 1 -(ni/nt)$, where *ni* is the number of OTUs observed exactly once, and *nt* is the total number of sequences.

Figure legends

Figure 1 Phylogenetic reconstruction of bacterial and archaeal 16S rRNA gene clones in the library derived from Carrizo sediment. Percentages of bacterial phylogenetic lineages detected in 16S rRNA gene clone library based on OPUs and the composition of the major groups (*Delta-*, *Beta-* and *Gammaproteobacteria*) and *Euryarchaeota* are shown in detail.

Figure 2 Proposed sulfur-metabolizing profile of the Carrizo community based on BLAST hits of protein homologues found in the metagenome data. The number of putative genes encoding for each particular enzyme class involved in the potential transformation of each molecule is specifically shown in brackets.

Figure 3 Proposed nitrogen-metabolizing profile of the Carrizo community based on BLAST hits of protein homologues found in the metagenome data. The number of putative genes encoding for each particular enzyme class involved in the potential transformation of each molecule is specifically shown in brackets.

Figure 4 Genomic content of cLDC0376 **(A)** and cLDC0001 **(B)** contigs. The GC-content of the contig is plotted with a window of 16,889 and 20,014 nucleotides, respectively. **(A)** As shown, the genes of cLDC0376 are organized in a tight cluster preceded by a phage integrase and three transposases. The location of genes with similar genome arrangements as *Sulfurovum* sp. NBC37-1 and *Sulfurospirillum deleyianum* DSM 6946 are shown. **(B)** cLDC0001 exemplified the horizontal transfer of a nitrate assimilation gene cluster (green). The GC percentage is indicated as a blue (low) to black (medium) and red (high) gradient.