



## UvA-DARE (Digital Academic Repository)

### Metagenomic Insights into the Uncultured Diversity and Physiology of Microbes in Four Hypersaline Soda Lake Brines

Vavourakis, C.D.; Ghai, R.; Rodriguez-Valera, F.; Sorokin, D.Y.; Tringe, S.G.; Hugenholtz, P.; Muyzer, G.

**DOI**

[10.3389/fmicb.2016.00211](https://doi.org/10.3389/fmicb.2016.00211)

**Publication date**

2016

**Document Version**

Final published version

**Published in**

Frontiers in Microbiology

**License**

CC BY

[Link to publication](#)

**Citation for published version (APA):**

Vavourakis, C. D., Ghai, R., Rodriguez-Valera, F., Sorokin, D. Y., Tringe, S. G., Hugenholtz, P., & Muyzer, G. (2016). Metagenomic Insights into the Uncultured Diversity and Physiology of Microbes in Four Hypersaline Soda Lake Brines. *Frontiers in Microbiology*, 7, [211]. <https://doi.org/10.3389/fmicb.2016.00211>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*



# Metagenomic Insights into the Uncultured Diversity and Physiology of Microbes in Four Hypersaline Soda Lake Brines

Charlotte D. Vavourakis<sup>1</sup>, Rohit Ghai<sup>2,3</sup>, Francisco Rodriguez-Valera<sup>2</sup>, Dmitry Y. Sorokin<sup>4,5</sup>, Susannah G. Tringe<sup>6</sup>, Philip Hugenholtz<sup>7</sup> and Gerard Muyzer<sup>1\*</sup>

<sup>1</sup> Microbial Systems Ecology, Department of Aquatic Microbiology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands, <sup>2</sup> Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, San Juan de Alicante, Spain, <sup>3</sup> Department of Aquatic Microbial Ecology, Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, České Budějovice, Czech Republic, <sup>4</sup> Research Centre of Biotechnology, Winogradsky Institute of Microbiology, Russian Academy of Sciences, Moscow, Russia, <sup>5</sup> Department of Biotechnology, Delft University of Technology, Delft, Netherlands, <sup>6</sup> The Department of Energy Joint Genome Institute, Walnut Creek, CA, USA, <sup>7</sup> Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences and Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia

## OPEN ACCESS

### Edited by:

Andreas Teske,  
University of North Carolina at Chapel  
Hill, USA

### Reviewed by:

Jinjun Kan,  
Stroud Water Research Center, USA  
Xi-Ying Zhang,  
Shandong University, China

### \*Correspondence:

Gerard Muyzer  
g.muyzer@uva.nl

### Specialty section:

This article was submitted to  
Extreme Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 02 December 2015

**Accepted:** 08 February 2016

**Published:** 25 February 2016

### Citation:

Vavourakis CD, Ghai R,  
Rodríguez-Valera F, Sorokin DY,  
Tringe SG, Hugenholtz P and  
Muyzer G (2016) Metagenomic  
Insights into the Uncultured Diversity  
and Physiology of Microbes in Four  
Hypersaline Soda Lake Brines.  
*Front. Microbiol.* 7:211.  
doi: 10.3389/fmicb.2016.00211

Soda lakes are salt lakes with a naturally alkaline pH due to evaporative concentration of sodium carbonates in the absence of major divalent cations. Hypersaline soda brines harbor microbial communities with a high species- and strain-level archaeal diversity and a large proportion of still uncultured poly-extremophiles compared to neutral brines of similar salinities. We present the first “metagenomic snapshots” of microbial communities thriving in the brines of four shallow soda lakes from the Kulunda Steppe (Altai, Russia) covering a salinity range from 170 to 400 g/L. Both amplicon sequencing of 16S rRNA fragments and direct metagenomic sequencing showed that the top-level taxa abundance was linked to the ambient salinity: *Bacteroidetes*, *Alpha*-, and *Gamma*-*proteobacteria* were dominant below a salinity of 250 g/L, *Euryarchaeota* at higher salinities. Within these taxa, amplicon sequences related to *Halorubrum*, *Natrinema*, *Gracilimonas*, purple non-sulfur bacteria (*Rhizobiales*, *Rhodobacter*, and *Rhodobaca*) and chemolithotrophic sulfur oxidizers (*Thioalkalivibrio*) were highly abundant. Twenty-four draft population genomes from novel members and ecotypes within the *Nanohaloarchaea*, *Halobacteria*, and *Bacteroidetes* were reconstructed to explore their metabolic features, environmental abundance and strategies for osmotic adaptation. The *Halobacteria*- and *Bacteroidetes*-related draft genomes belong to putative aerobic heterotrophs, likely with the capacity to ferment sugars in the absence of oxygen. Members from both taxonomic groups are likely involved in primary organic carbon degradation, since some of the reconstructed genomes encode the ability to hydrolyze recalcitrant substrates, such as cellulose and chitin. Putative sodium-pumping rhodopsins were found in both a *Flavobacteriaceae*- and a *Chitinophagaceae*-related draft genome. The predicted proteomes of both the latter and a *Rhodothermaceae*-related draft genome were indicative of a “salt-in” strategy of osmotic adaptation. The primary catabolic and respiratory pathways shared

among all available reference genomes of *Nanohaloarchaea* and our novel genome reconstructions remain incomplete, but point to a primarily fermentative lifestyle. Encoded xenorhodopsins found in most drafts suggest that light plays an important role in the ecology of *Nanohaloarchaea*. Putative encoded halolysins and laccase-like oxidases might indicate the potential for extracellular degradation of proteins and peptides, and phenolic or aromatic compounds.

**Keywords:** soda lake brines, *Nanohaloarchaea*, *Halobacteria*, *Bacteroidetes*, hydrolytics, cellulase, chitinase, rhodopsin

## INTRODUCTION

Modern soda lakes are characterized by high salt concentrations that can reach up to saturation and a stable high pH, typically between 9.5 and 11 (Jones et al., 1977; Tindall, 1988; Banciu and Sorokin, 2013). They occur in (semi-)arid regions world-wide where they form naturally through evaporative concentration. Their origin is not related to seawater (“athalassic”), but to seepage of groundwater that is rich in sodium ions ( $\text{Na}^+$ ) and  $\text{CO}_2$ , but low in calcium ( $\text{Ca}^{2+}$ ), and magnesium ( $\text{Mg}^{2+}$ ) ions that would otherwise precipitate the dissolved carbonates. Whereas in the water column of “thalassic” environments and inland salt lakes, the dominant soluble salt is sodium chloride, sodium carbonate/bicarbonate salts dominate in soda lake brines and create a naturally buffered haloalkaline system. Moderately saline soda lakes are believed to be amongst the most productive aquatic environments on Earth (Melack, 1981). Complete biogeochemical carbon, nitrogen and sulfur cycles can still occur in hypersaline (salinity  $>5\%$  w/v) soda lakes and become only partially hampered at salt-saturating conditions (Sorokin et al., 2014, 2015a, and references therein).

For hypersaline thalassic environments with circum-neutral pH, it is well-established from studies on solar salterns that salinity is the major environmental factor shaping planktonic prokaryotic community composition (e.g., Rodriguez-Valera et al., 1985; Oren, 1994; Casamayor et al., 2000, 2002; Benlloch et al., 2002; Ghai et al., 2011; Dillon et al., 2013). Seawater is led into an interconnected pond system where salinity gradually increases by evaporative concentration until sea salt can be harvested. Along the pond system, proteobacterial planktonic diversity decreases in favor of extreme halophilic *Archaea*, such as *Haloquadratum* sp. belonging to the class *Halobacteria* (phylum *Euryarchaeota*) and members of the candidate class *Nanohaloarchaea* (phylum *Nanohaloarchaeota*; Ghai et al., 2011; Narasingarao et al., 2012; Martínez-García et al., 2014), as well as some moderate and extreme halophilic *Bacteria* such as the *Gammaproteobacteria* *Spiribacter salinus* (López-Pérez et al., 2013; León et al., 2014), *Halovibrio* and *Halospina* (Sorokin et al., 2006), and *Salinibacter ruber* in the *Bacteroidetes* phylum (Antón et al., 2000, 2002; Mongodin et al., 2005).

A similar influence of prevailing salt concentrations on microbial community composition could be inferred from previous culture-independent studies of alkaline soda lake brines. Various bacterial genera belonging to the *Alphaproteobacteria*, *Firmicutes*, *Gammaproteobacteria*,

*Bacteroidetes*, and *Cyanobacteria* are found in soda brines with salinities up to 250 g/L (Humayoun et al., 2003; Mesbah et al., 2007; Dimitriu et al., 2008; Pagaling et al., 2009; Asao et al., 2011; Lanzen et al., 2013), while above these salinities *Euryarchaeota* belonging to the *Halobacteria* and the *Methanosarcinales* are found to dominate in natural lakes (Grant et al., 1999; Ochsenreiter et al., 2002; Mesbah et al., 2007; reviewed by Sorokin et al., 2014, 2015a) and in artificial soda concentration ponds (Gareeb and Setati, 2009; Simachew et al., 2015).

Most culture-independent studies performed in these environments world-wide are in agreement that soda lakes harbor a surprisingly high 16S rRNA diversity and abundance of still uncharacterized microbes. At high salinities, the high diversity within several genera and species of the class *Halobacteria* is not in proportion to the relatively low number of soda lake isolates obtained from this group of mostly aerobic heterotrophs. Only recently the first alkaliphilic members of the *Halobacteria* capable of growing on cellulose and chitin were isolated from various soda lakes, showing that these organisms can fulfill an environmental role as primary organic matter degraders (Sorokin et al., 2015d). It is well-established, and has received considerable interest from the biotechnology-industry, that many members of the *Halobacteria* are able to degrade proteins and peptides via extra-cellular proteases, also referred to as “halolysins” (De Castro et al., 2006). Almost all of these halolysins are subtilisin-like serine proteases (peptidase subfamily S8A) and quite a few are isolated, including from alkaliphilic members of the *Halobacteria*. Many members of the phylum *Bacteroidetes* are also able to consume biopolymers, such as chitin, cellulose, xylan and proteins, and are therefore often viewed as a key group for organic matter remineralization in the marine environment (e.g., Bauer et al., 2006). Interestingly, with the exception of *Natronoflexus pectinovorans* (phylum *Bacteroidetes*), polysaccharide-degrading *Bacteria* isolated from soda lakes belong mostly to the *Firmicutes* (genera *Bacillus*, *Clostridium*), *Actinobacteria* (genera *Cellulomonas* and *Dietzia*), *Gammaproteobacteria* (*Alkalimonas*, *Marinimicrobium*), and *Fibrobacteres* (genus *Chitinovibrio*; Sorokin et al., 2014, 2015a).

With the advent of next generation sequencing, direct metagenomic approaches have been successfully applied to extreme environments harboring relatively low-diversity communities. They have proven very useful to construct snapshots of the overall microbial community structure and to aid in the reconstruction of draft or even complete genomes of uncultured microorganisms (e.g., Tyson et al., 2004; Ghai

et al., 2011; Narasingarao et al., 2012; Sharon and Banfield, 2013; Fernández et al., 2014a,b). Notably, nanohaloarchaeal 16S rRNA sequences were first recovered from the hypersaline soda lake Magadi in the east-African Rift Valley (Grant et al., 1999) and since then four near-complete genomes belonging to the “*Candidatus Nanohaloarchaea*” have been reconstructed from metagenomes of various hypersaline aquatic habitats (Ghai et al., 2011; Narasingarao et al., 2012; Martínez-García et al., 2014). Based on 16S rRNA phylogeny of sequences recovered from hypersaline and haloalkaline habitats worldwide, *Nanohaloarchaea* were first viewed as a deep-branching sister-group of the class *Halobacteria* within the phylum *Euryarchaeota* (Narasingarao et al., 2012; Oren, 2014). Based on the recent genome reconstructions however, a new candidate phylum “*Nanohaloarchaeota*” within a “DPANN” super-phylum was proposed (Rinke et al., 2013; Castelle et al., 2015), but this view remains controversial since the outcome of archaeal phylogenetic analyses appears to depend on the choice of marker genes and reference genomes, among other factors (Petitjean et al., 2015). In addition to the absence of cultured isolates from this class and their disputed phylogeny, the lifestyle and ecological function of *Nanohaloarchaea* remains enigmatic.

Here we determined for the first time the overall structure of microbial communities from the brines of three natural hypersaline soda lakes and a trona crystallizer located in the Kulunda Steppe (Altai, Russia), covering a salinity range between 170 and 400 g/L with a combination of both 16S rRNA amplicon and direct metagenome sequence analyses. In the past 15 years, these lakes have been intensively studied by classical microbiology approaches providing a good background for the interpretation of the metagenomics results. We reconstructed several draft genomes of yet uncultured populations within the phylum *Bacteroidetes*, class *Halobacteria*, and candidate class *Nanohaloarchaea* in order to explore their metabolic potential with respect to organic carbon remineralization in soda brines and other putative roles in biogeochemical cycling.

## MATERIALS AND METHODS

### Site Description, Sample Collection and DNA Extraction

The four sampled soda lakes are located in the southern part of the Kulunda Steppe, South-Western Siberia (Altai region), Russia. The exact soda lake locations, chemical parameters and general appearance of the brine at the time of sampling are given in **Table 1**. The eutrophic lakes Bitter-1 (B1), Tanatar trona crystallizer (Tc), and Tanatar-5 (T5) are part of the Bitter and Tanatar systems, respectively, and were formed on tertiary depositions of trona-cemented sands (Isachenko, 1951). The oligotrophic Picturesque Lake (PL) is located in the depression of the Salt Lake Steppe (sediments - gray kaolinite clays). All lakes are hypersaline, shallow ponds subjected to a very unstable water regime and with temperature fluctuations from +35 to -35°C (Foti et al., 2008). At the time of sampling the salt concentration ranged from 170 to 400 g/L sodium carbonates/chloride and a pH from 9.5 to 10.3 in different lakes.

Brine samples were collected in sterile glass bottles in July 2010 and 2011 and prefiltered (Whatman “White Ribbon” paper filters) to remove suspended sediment particles and aggregates. Cells from 1 L brine were pelleted by high speed centrifugation (30 min at 13,000 rpm with a Sorvall SLC-1500 rotor) and resuspended in 2-4 M NaCl. Smaller volumes were finally pelleted (10 min at 20,817 × g) in 2 mL Eppendorf tubes and stored at -70°C until further processing. High molecular weight DNA was extracted with the classical phenol/chloroform method (Marmur, 1961) and isopropanol precipitation after resuspending the pellets in 10 mL 0.1 M Tris (pH 8)/10 mM EDTA and pretreatment with lysozyme, SDS, proteinase K, and RNase. The quality and quantity of the extracted DNA was checked by agarose gel electrophoresis using standard markers of known concentration.

### Amplicon Sequencing and Analysis

The V6-V8 region of 16S rRNA was amplified using primers 926F (5'- cct atc ccc tgt gtg cct tgg cag tct cag AAA CTY AAA KGA ATT GRC GG- 3') and 1392R (5' - cca tct cat ccc tgc gtg tct ccg act cag - <XXXXX> - ACG GGC GGT GTG TRC - 3') universal to both *Archaea* and *Bacteria* (Tremblay et al., 2015). Primer sequences were modified by the addition of 454 A or B adapter sequences. In addition, the reverse primer included a 5 bp barcode for multiplexing of samples during sequencing. Twenty-microlitre PCR reactions were performed in duplicate and pooled to minimize PCR bias using 0.4 µl Advantage GC 2 Polymerase Mix (Advantage-2 GC PCR Kit, Clontech), 4 µl 5X GC PCR buffer, 2 µl 5 M GC Melt Solution, 0.4 µl 10 mM dNTP mix (MBI Fermentas), 1.0 µl of each 25 nM primer, and 10 ng sample DNA. The thermal cycler protocol was 95°C for 3 min, 25 cycles of 95°C for 30 s, 50°C for 45 s, and 68°C for 90 s, and a final 10 min extension at 68°C. PCR amplicons were purified using SPRI Beads (Beckman Coulter) and quantified using a Qubit fluorometer (Invitrogen). PCR products were diluted to 10 ng/µl and mixed in equal concentrations. Emulsion PCR and sequencing of the PCR amplicons were performed following the Roche 454 GS FLX Titanium technology manufacturer's instructions. Sequencing tags were analyzed using the software tools PyroTagger (<http://pyrotagger.jgi-psf.org>) and MacQIIME (<http://www.wernerlab.org/software/>).

### Metagenome Sequencing and Raw Read Analysis

Environmental DNA (500 ng) was sheared using the Covaris E210 and size selected for 270 bp fragments using Agencourt Ampure Beads (Beckman Coulter) prior to library construction with the KAPA Illumina library creation kit (KAPA Biosystems). Each library was quantified by qPCR and sequenced on one lane of the Illumina HiSeq 2000 platform to generate paired-end 150 bp reads. Low quality regions of the paired-end raw reads were trimmed using DynamicTrim (Cox et al., 2010). Top-level taxonomic profiles of the metagenomes were determined according to Ghai et al. (2011). Candidate 16S rRNA encoded metagenomic reads were initially selected by comparing a subsample of 2 million sequence pairs with a small manually-curated database. Reads with a minimum alignment length of

**TABLE 1 | General features of the sampled soda lakes, the obtained metagenomic raw sequence reads, and the assembled contigs.**

Lake	Lake Tanatar-5 (T5)	Picturesque Lake (PL)	Lake Tanatar trona crystallizer (Tc)	Lake Bitter-1 (B1)
Location	51.62 N 79.84 E	51.73 N 79.87 E	51.65 N 79.75 E	51.67 N 79.91 E
Sampling year	2010	2010	2011	2010
Salinity (g/L)	170	250	300	400
pH brine	9.9	9.5	9.9	10.2
Na <sub>2</sub> CO <sub>3</sub> alkalinity (M)	0.8	1.1	4.0	2.0
Total carbonate alkalinity (M)	1.9	2.8	5.0	4.4
Brine appearance	Muddy	Muddy	Orange-red	Oily yellow
Metagenome ID	T5-Br10	PL-Br10	Tc-Br11	B1-Br10
SRA Accession	INGZ-SRA052017	INHF-SRA052008	INHA-SRA052024	INHB-SRA052018
Number of read pairs	206,674,382	352,665,522	191,945,135	411,685,143
Estimated mean coverage	140x	190x	129x	202x
Number of assembled contigs	636,016	1,140,049	458,620	482,919
Number of contigs ≥5 kb	19,350	25,098	9,426	15,551
GenBank Accession (contigs ≥5 kb)	LFIK000000000	LKMJ000000000	LFFM000000000	LFCJ000000000
Average contig length (b)	1,005	974	895	1,115
Maximum contig length (b)	204,877	192,059	134,932	155,378
Total assembly length (b)	639,293,147	1,110,081,880	410,404,879	538,272,225

Raw sequence reads have been deposited to the NCBI Sequence Read Archive (SRA) and contigs larger than 5 kb at DDBJ/EMBL/GenBank.

90 bases (min identity 80%, *e*-value 1e-5) were extracted and compared to sequences in the Ribosomal Database Project (Cole et al., 2014). The best-named hit was used to classify the reads to top level taxa. Chloroplast 16S rRNA was manually filtered out and taxa containing less than 1% of the total assigned 16S rRNA reads were not considered.

## Reconstruction and Analysis of Draft Genomes

Assembly of the metagenomic reads was performed with MEGAHIT v1.0.1-13-gd6932f9 (Li et al., 2015). Protein coding genes on contigs longer than 5kb were predicted with Prodigal (Hyatt et al., 2010), tRNAs with tRNAscan (Lowe and Eddy, 1997), rRNAs using meta\_rna (Huang et al., 2009) and annotated using BLAST against the non-redundant database Pfam (Bateman et al., 2004), COGs (Tatusov et al., 2003), and TIGRFams (Haft et al., 2001). Genes at the end of contigs and incomplete genes without valid start/stop codons were removed. Based on sequence similarity against the non-redundant NCBI database, the best hit for each gene was determined and used to bin to top level taxa. Only those contigs that contained a minimum of three genes and had not more than half of the genes hitting different taxa were retained. Further binning of the contigs belonging to *Bacteroidetes*, *Nanohaloarchaea*, or *Halobacteria* was done by differential coverage binning with CONCOCT (Alneberg et al., 2014) after using subsets of 20 million sequence reads of each dataset to estimate coverage by mapping reads with min 95% nucleotide identity and 50 bp overlap. Possible bin contamination and strain heterogeneity were estimated with CheckM v0.9.7 (Parks et al., 2014) and only those with reasonable concatenated length (>0.5–0.8 Mb) and with contamination below 5% were retained for subsequent analysis.

Taxonomic affiliation of the selected bins was determined via phylogenetic trees based on concatenates of all proteins identified within Clusters of Orthologous Groups (COG; Tatusov et al., 2003) and shared with complete or near-complete reference genomes (RefSeq NCBI, plasmids excluded), since no 16S rRNA genes were found on the contigs. Shared proteins were concatenated and aligned using Kalign (Lassmann and Sonnhammer, 2005) and a maximum-likelihood tree was made using the software program MEGA version 6 (Tamura et al., 2013). The completeness of the reconstructed bacterial and archaeal genomes was assessed based on the presence of 112 essential genes (Albertsen et al., 2013) and 53 core genes (Narasimgarao et al., 2012) respectively, using HMMER. Using an *e*-value cut-off of 1e-5, lower and upper bound predictions of completeness were obtained using all predicted genes or only the best annotations (hcov and qcov >80%). Average nucleotide identity (ANI) and conserved DNA fraction between reconstructed and/or reference genomes were calculated based on the whole genome sequence as described by Goris et al. (2007).

The carbon-, nitrogen- and sulfur metabolic potential of the draft genomes was inferred from all proteins identified within COG, as well as from KEGG pathways and BRITE hierarchies after re-annotating the predicted proteins from the standard draft genomes automatically using KAAS (KEGG Automatic Annotation Server: <http://www.genome.jp/kegg/kaas/>; Moriya et al., 2007). Additionally, assigned K-numbers were used to screen for the presence of functional marker genes (Lauro et al., 2011; Llorens-Marès et al., 2015). Putative encoded carbohydrate-active enzymes in the draft genomes were identified by sequenced-based annotation of the predicted proteins from all coding sequences using the CAZymes Analysis Toolkit (CAT; Park et al., 2010; Lombard et al., 2014).

To determine the general mode of osmotic adaptation of the putative microorganisms for which we reconstructed the genomes, the isoelectric points (pI) of the coding sequences with a minimum length of 30 amino acids were calculated using the program IEP in the EMBOSS package (Rice et al., 2000). The pI distribution (bin width 0.1) for each composite genome was compared to the pI distribution of selected, complete reference genomes available in RefSeq. Genes likely to be involved in osmotic adaptation were identified through their annotations and best-hits (see above), namely those involved in osmolyte synthesis and uptake or those encoding putative ion transporters and channels.

## Read Recruitments

Population abundance in the different soda lake brines was estimated via read recruitments to the draft genomes from subsamples of 20 million reads from the different metagenomic datasets. A cut-off of 95% identity in at least 50 bases was used and the number of hits was normalized against the size of the genomes and metagenomes and expressed as Reads Per Kilobase of sequenced reads per Gigabase of mapped reads (RPKG; López-Pérez et al., 2013).

## Data Availability

The raw sequence reads of the four metagenomics datasets have been deposited to the NCBI Sequence Read Archive (SRA; Accession Numbers given in **Table 1**). The assembled contigs larger than 5 kb have been deposited as Whole Genome Shotgun projects at DDBJ/EMBL/GenBank (Accession Numbers given in **Table 1**). The reconstructed draft genomes have been deposited as Whole Genome Shotgun projects at DDBJ/EMBL/GenBank under the accessions LKMK00000000-LKNH00000000 (complete list given in Supplementary Table 2). All versions described in this paper are version XXXX010 00000.

## RESULTS AND DISCUSSION

### Microbial Community Composition Based on 16S rRNA Gene Sequences

#### Top Level Taxa Abundance in Relation to Salinity

Microbial community composition of the different soda lake brines was compared based on metagenomic reads containing fragments of the 16S rRNA gene as well as on OTUs assigned to the amplicon sequences (**Figure 1**). The majority of the selected metagenomic reads originating from the saturated salt brines (total salinity  $\geq 250$  g/L) of Bitter-1 (B1), Tanatar trona crystallizer (Tc), and Picturesque Lake (PL) belonged to the archaeal phylum *Euryarchaeota* (>91%). Those from Tanatar-5 (T5) with a total salinity of 170 g/L belonged to the bacterial taxa *Bacteroidetes* (34%), *Gammaproteobacteria* (33%), and *Alphaproteobacteria* (12%) and to a lesser extent to the *Actinobacteria*, *Firmicutes*, *Cyanobacteria*, and unclassified bacterial groups (<5%). *Euryarchaeota* still constituted a considerable fraction of the prokaryotic community in the brine of T5 (7%). Similarly, 16S rRNA gene amplicon sequences belonging to *Euryarchaeota* dominated in the three most saline

brine datasets, while those belonging to *Bacteria* dominated the lower salinity dataset. Our results are in agreement with previous culture-independent studies that reported the dominance of *Euryarchaeota* in 16S rRNA gene clone libraries from hypersaline soda lakes (Grant et al., 1999; Ochsenreiter et al., 2002; Mesbah et al., 2007; Simachew et al., 2015). Moreover, all bacterial top level taxa that were previously found to be common in soda lakes world-wide with a maximum salinity of 250 g/L (Humayoun et al., 2003; Mesbah et al., 2007; Dimitriu et al., 2008; Pagaling et al., 2009; Asao et al., 2011; Lanzen et al., 2013) appeared abundant (>1%) in both our metagenomic and amplicon T5 dataset.

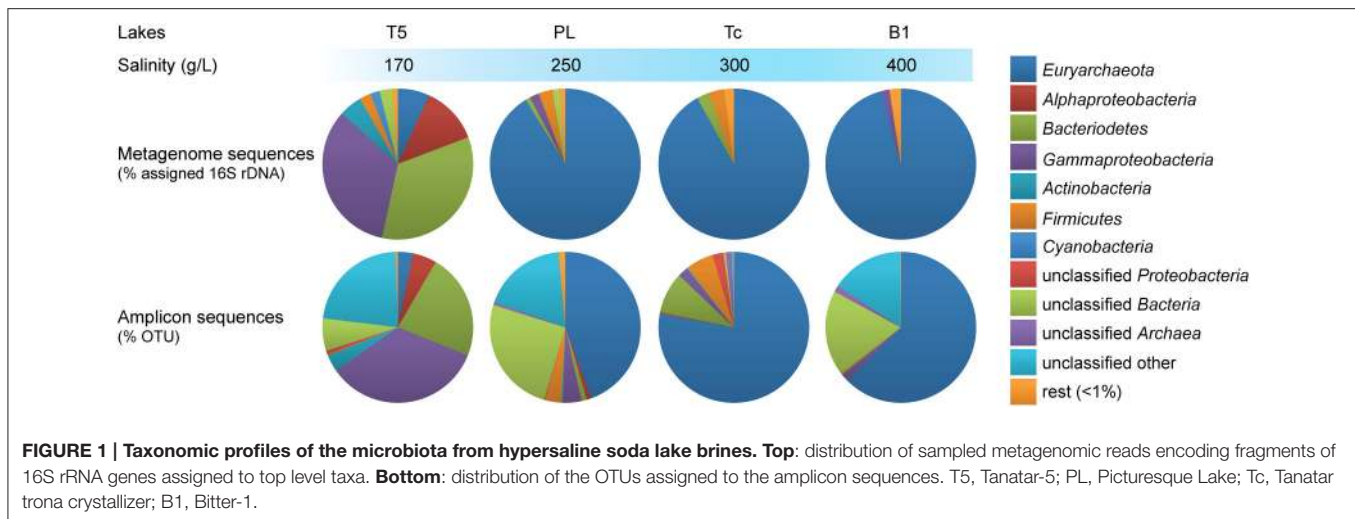
Amplicon and shotgun sequence-based taxonomic profiling introduced different biases. While the direct metagenomic approach might have left rare taxa undetected (Lynch and Neufeld, 2015), amplification biases in complex DNA mixtures might have distorted the original 16S rRNA gene ratios. Most evidently for the three most saline lakes, rare (here <1%) bacterial taxa were overrepresented in the amplicon datasets compared to the metagenomes, while archaeal 16S sequences were underrepresented. Furthermore, between 16 and 23% of the 16S amplicon sequences from the brines of B1, PL and T5 could not be assigned to any top level bacterial or archaeal taxon (“unclassified”). Since no “unclassified” sequences were found in taxonomic profiles derived from the metagenomes, these are suspected to originate from eukaryote organelles other than chloroplasts. *Cyanobacteria* were absent in the three most saline brines when considering the directly sequenced metagenomes, but present in the same brines when considering the amplicon sequences. Finally, on the phylum and class level, the total number of detected taxa (including “Unclassified” *Bacteria* and *Archaea*) in the metagenomes was inversely related to salinity with six taxa detected for B1 (400 g/L salt) and 14 for T5 (Supplementary Table 3). This trend was not apparent in the amplicon sequencing datasets where the brines of PL and T5 appeared the most and the least rich (19 and 11 top level taxa OTUs), respectively.

#### Cultured vs. Unknown Members of the most Abundant Top Level Taxa

While the directly sequenced, short-read metagenomes yielded in our opinion more realistic snapshots of the relative 16S rRNA gene abundances of the high-level community constituents, we considered only the longer amplicon sequences (~450 nt vs. <150 nt) for further phylogenetic assignment down to the genus-level, focusing on the identified abundant top-level taxa (**Table 2**).

#### *Euryarchaeota*

More than 99% of the OTUs assigned to *Euryarchaeota* in the three amplicon sequencing datasets from the three most saline brines of PL, Tc, and B1, were affiliated with 11 distinct and known genera from the class *Halobacteria*. Among them, *Halorubrum*- and *Natrinema*-related sequences were the most abundant ones. The remaining largest fraction of the OTUs, as well as all *Halobacteria*-related sequences from the brine of T5, could not be assigned to any known genus.



Few *Methanosarcinaceae*-related sequences of the genus *Methanolobus* (class “*Methanomicrobia*”) were found in the brine of T5 (1.2% of *Euryarchaeota*-related OTUs or 0.04% of the total OTUs—see **Table 2**). The presence of this specific genus of moderately salt-tolerant, methylotrophic methanogens was also detected in 16S clone libraries obtained from the hypersaline lake Fazda (Wadi an Natron, Egypt; Mesbah et al., 2007). More recently, *Methanolobus* was also found in the sediments of hypersaline soda lakes of the Kulunda Steppe by a functional gene assay (Sorokin et al., 2015b). In addition, two *Methanolobus* isolates were obtained and potential methanogenic activity up to soda saturating conditions confirmed. Although methanogens are typically sediment-affiliated and suspended clay particles might have passed the pre-filtration step in this study, oxygen solubility in hypersaline brines is low (e.g., Grant, 2004) and therefore *Methanolobus*-related sequences might also originate from methanogens thriving in the brine of T5 under anoxic conditions.

### *Bacteroidetes*

Most of the *Bacteroidetes* OTUs in T5 belonged to the genus *Gracilimonas* (~83%). This genus comprises a group of marine (Choi et al., 2009; Cho et al., 2013) and salt mine (Wang et al., 2013) isolates and related 16S gene fragments were recovered recently from a haloalkaline soil (Keshri et al., 2013). In the datasets of the more saline brines of B1, Tc, and PL, *Gracilimonas*-related sequences were also the dominant *Bacteroidetes*. The two important remaining *Bacteroidetes* fractions in the brine of T5 were affiliated with unknown genera from the families *Chitinophagaceae* (~11%) and *Flavobacteriaceae* (~5%).

### *Alphaproteobacteria*

From the OTUs assigned to the *Alphaproteobacteria* in the T5 dataset, the dominant fraction (~63%) was affiliated with the family *Rhodobacteraceae* (order *Rhodobacterales*) and belonged to the genera *Roseinonatronobacter*, *Rhodobacter* and *Rhodobaca*. The genus *Roseinonatronobacter* comprises the haloalkaliphilic members of a unique group of aerobic, bacteriochlorophyll

*a*-containing, sulfur-oxidizing lithoheterotrophs that were first isolated from soda lake microbial mats (Sorokin et al., 2000) and later from the low-salt epilimnion of the alkaline Mono Lake (Boldareva et al., 2007). Haloalkaliphilic members of the purple non-sulfur bacteria belonging to the latter two genera are frequently detected in soda lakes (Rees et al., 2004; Kompantseva et al., 2007; Asao et al., 2011). While members of the genus *Rhodobaca* were isolated from soda lakes (Milford et al., 2000; Boldareva et al., 2008), so far no alkaliphilic *Rhodobacter* isolates are obtained. Most of the *Rhodobacteraceae*-related OTUs from T5 could not be assigned to a known genus, which was also a conclusion from previous research on Kulunda soda lakes that focused on this group (Kompantseva et al., 2010). Another large alphaproteobacterial fraction (~32%) was assigned to the order *Rhizobiales*, but no family could be assigned. The presence of the rhizobial genus *Mesorhizobium* was reported previously in the sediment and dry mud from a Kenyan soda lake (Mwirichia et al., 2011) and in a recent molecular survey of Kulunda soda lake sediments (Tourova et al., 2014). Although there is a high probability that the *Rhizobiales*-related sequences in this study are from terrestrial origin, the order includes, besides plant rhizobia, plant- and animal-pathogens, photosynthetic purple non-sulfur bacteria that occur in soda lakes, such as the genera *Rhodospseudomonas*, *Rhodoblastus*, and *Rhodomicrobium* (Tourova et al., 2011). In the datasets of the more moderate saline brines of PL and T5, a substantial fraction of the alphaproteobacterial OTUs affiliated with the *Rhodospirillaceae*, another phylogenetic group of purple non-sulfur bacteria (Imhoff, 1995), still without soda lake isolates.

### *Gammaproteobacteria*

From the OTUs assigned to the *Gammaproteobacteria* in the T5 dataset, the dominant fraction (~59%) was affiliated with no known order; the largest remaining fractions belonged to the genera *Thioalkalivibrio* (order *Chromatiales*; ~37%) and *Halomonas* (order *Oceanospirillales*; ~2%), comprising haloalkaliphilic sulfur-oxidizing bacteria and aerobic heterotrophs commonly occurring in soda lakes with various salinities

**TABLE 2 | Genus assignments of the detected OTUs from 16S rRNA amplicon sequences belonging to the most abundant top level taxa from different soda lake brines.**

% Assigned of total OTUs		% Assigned of total OTUs-continued				
		B1	Tc	PL	T5	
<b>Euryarchaeota</b>						
<i>Halobacteriales<sup>a</sup>/Halobacteriaceae<sup>a</sup>/Halorubrum</i>		23.5	23.7	5.8	0.00	
<i>Halobacteriales<sup>b</sup>/Halobacteriaceae<sup>b</sup>/Natrinema</i>		0.94	1.69	0.35	0.00	
<i>Halobacteriales<sup>c</sup>/Halobacteriaceae<sup>c</sup>/other<sup>d</sup></i>		0.11	0.69	0.59	0.00	
<i>Halobacteriales<sup>c</sup>/Halobacteriaceae<sup>c</sup>/none</i>		38.2	51.7	37.7	3.2	
<i>Methanosarcinales/Methanosarcinaceae/Methanobubis</i>		0.00	0.00	0.00	0.04	
<i>Other<sup>e,f</sup></i>		0.03	0.20	0.03	0.00	
<b>Bacteroidetes</b>						
<i>Bacteroidales/Marinilabiaceae/none</i>						
<i>Flavobacteriales/Flavobacteriaceae/none</i>		0.00	0.00	0.01	1.05	
<i>Flavobacteriales/other<sup>g</sup></i>		0.00	0.00	0.01	0.00	
<i>Sphingobacteriales/Chitinophagaceae/Balneola</i>		0.01	0.00	0.00	0.00	
<i>Sphingobacteriales/Chitinophagaceae/Graclimonas</i>		0.02	8.3	0.72	18.8	
<i>Sphingobacteriales/Chitinophagaceae/none</i>		0.00	0.22	0.25	2.50	
<i>Sphingobacteriales/other<sup>h,i</sup></i>		0.00	0.00	0.004	0.00	
<i>Sphingobacteriales/Rhodothermaceae/none</i>		0.00	0.24	0.00	0.00	
<i>Sphingobacteriales/Rhodothermaceae/Salsaeta</i>		0.00	0.06	0.00	0.00	
<i>Other<sup>h,j</sup></i>		0.01	0.01	0.18	0.24	
<b>Alphaproteobacteria</b>						
<i>Rhizobiales/other<sup>g,k</sup></i>		0.00	0.02	0.21	1.65	
<i>Rhodobacteriales/Rhodobacteraceae/Rhodobaca</i>		0.00	0.00	0.00	0.04	
<i>Rhodobacteriales/Rhodobacteraceae/Rhodobacter</i>		0.00	0.00	0.01	0.28	
<i>Rhodobacteriales/Rhodobacteraceae/Roseinatronobacter</i>		0.01	0.00	0.09	1.13	
<i>Rhodobacteriales/Rhodobacteraceae/none</i>		0.14	0.28	0.45	1.81	
<i>Rhodospirillales/other<sup>g</sup></i>		0.00	0.00	0.01	0.00	
<i>Rhodospirillales/Rhodospirillaceae/none</i>		0.00	0.00	0.05	0.24	
<i>Other<sup>h</sup></i>		0.00	0.00	0.07	0.00	
<b>% Assigned of total OTUs</b>						
<b>Gammmaproteobacteria</b>						
<i>Alteromonadales/Alteromonadaceae/Hallela</i>			0.01	0.00	0.00	
<i>Alteromonadales/Idiomarinaceae/none</i>			0.01	0.00	0.01	
<i>Alteromonadales/other<sup>g</sup></i>			0.01	0.00	0.02	
<i>Chromatiales/Chromatiaceae/Halochromatium</i>			0.00	0.00	0.00	
<i>Chromatiales/Ectothiorhodospiraceae/Thioalkalivibrio</i>			0.07	0.26	0.71	
<i>Chromatiales/Ectothiorhodospiraceae/Aquissalmonas</i>			0.01	0.01	0.13	
<i>Chromatiales/Ectothiorhodospiraceae/Halorhodospira</i>			0.15	0.57	0.15	
<i>Chromatiales/Ectothiorhodospiraceae/other<sup>l</sup></i>			0.00	0.00	0.02	
<i>Chromatiales/Ectothiorhodospiraceae/none</i>			0.00	0.01	0.01	
<i>Chromatiales/Halotheobacillaceae/Thioalkalibacter</i>			0.00	0.00	0.01	
<i>Chromatiales/other<sup>g</sup></i>			0.00	0.00	0.01	
<i>Gammapr. incertae sedis/Methylotrum/none</i>			0.00	0.00	0.01	
<i>Legionellales/Legionellaceae/none</i>			0.00	0.00	0.002	
<i>Oceanospirillales/Hahellaceae/Hahella</i>			0.003	0.00	0.00	
<i>Oceanospirillales/Halomonadaceae/Halomonas</i>			0.86	0.01	0.57	
<i>Oceanospirillales/Halomonadaceae/none</i>			0.00	0.03	0.81	
<i>Oceanospirillales/Halomonadaceae/Salicola</i>			0.00	0.00	0.002	
<i>Oceanospirillales/other<sup>g</sup></i>			0.00	0.00	0.06	
<i>Thioarchaeales/Piscirickettsiaceae/Thioalkalimicrobium</i>			0.00	0.00	0.004	
<i>Xanthomonadales/Xanthomonadaceae/none</i>			0.00	0.004	0.00	
<i>Xanthomonadales/Xanthomonadaceae/Stenotrophomonas</i>			0.00	0.00	0.002	
<i>Other<sup>h</sup></i>			0.09	1.22	1.41	
<b>Remaining OTUs</b>			<b>35.9</b>	<b>10.8</b>	<b>49.5</b>	<b>34.7</b>

T5, Tanatar-5; PL, Picturesque Lake; Tc, Tanatar Irona crystallizer; B1, Bitter-1.

<sup>A</sup> Haloteracales/Haloteracaceae acc. to Gupta et al. (2014).

<sup>B</sup> Natribales/Natribaceae according to Gupta et al. (2014).

<sup>C</sup> Classification before Gupta et al. (2014).

<sup>D</sup> Natronorubrum, Halopiger, Haloplanus, Halostagnicola, Haloterrigena, Natribalba, Natronococcus, Natronolimnobius, Natronomonas.

<sup>E</sup> Thermoplasmatales.

<sup>F</sup> No class assigned.

<sup>G</sup> No family assigned.

<sup>H</sup> No order assigned.

<sup>I</sup> Family Saprospiraceae.

<sup>J</sup> Order Bacteroidales.

<sup>K</sup> Family Bradyrhizobiaceae.

<sup>L</sup> Alkallimnicola, Thioalkalispira, Thiohalospira.



(Sorokin et al., 2014). In contrast, from the OTUs assigned to the *Gammaproteobacteria* in the dataset B1, the soda lake with the highest salinity, the dominant fractions belonged to *Halomonas* (~71%) and purple sulfur bacteria (genus *Halorhodospira*, order *Chromatiales*; ~13%).

## Draft Genomes of Putative Predominant, Yet Still Uncultured Community Members

After assembling the metagenomic reads into contigs and classifying all  $\geq 5$  kb contigs to top-level taxa, we obtained several high-quality (<4% estimated contamination using CheckM) draft genomes of uncultured members of the *Euryarchaeota*, *Bacteroidetes* and the candidate class *Nanohaloarchaea* (Table 3), as well as a large number of contigs belonging to *Eukaryota*, *Gamma*- and *Alphaproteobacteria* and *Actinobacteria* (Supplementary Table 1). The obtained draft genomes represent “population” genomes i.e., not clonal (Sharon and Banfield, 2013) of single species or strains. Below we discuss their inferred phylogeny based on genome trees (Figures 2A–E) and average nucleotide identities (ANIs; Supplementary Tables 4–8) with known organisms, their metabolic potential and principal modes of osmotic adaptation (Figure 3).

### Candidate Class *Nanohaloarchaea*

#### Phylogeny

From B1-Br10 and PL-Br10 we obtained eight draft genomes belonging to low-GC Archaea of the candidate class *Nanohaloarchaea* (Table 3; Figure 2A; Supplementary Table 4). Based on the genome tree, four of these reconstructions, two from each brine, were most closely related to the draft genome of “*Candidatus Haloredivivus*” obtained from metagenomes in a solar saltern pond (Ghai et al., 2011). They belonged to four different genera based on average nucleotide identity (all ANIs <74.9%). The remaining four draft genomes belonged to one to three different genera, distantly related to “*Candidatus Nanosalina*” obtained from a hypersaline lake (Narasimgarao et al., 2012; all ANIs < 74.5%), while the drafts B1-Br10-U2g1 and PL-Br10-U2g19 belonged to the same species thriving at different brines (ANI = 99.4%).

#### *Xenorhodopsins and heterotrophy*

We found several genes that support the predominant heterotrophic lifestyle previously proposed for “*Candidatus Haloredivivus*” (Ghai et al., 2011; Narasimgarao et al., 2012). Notably, out of the eight draft genomes, six encoded putative xenorhodopsins for which the exact physiological role still remains enigmatic (Ugalde et al., 2011; Supplementary Figure 1, Supplementary Table 9). Starch-degrading metabolism was evident from encoded alpha-amylases and related glucoside hydrolases (CAZy families GH13 and GH57) found in seven out of eight drafts. In five drafts subtilisin-like proteins (peptidase S8) were encoded, suggesting degradation potential for extracellular proteins and peptides. Furthermore, in B1-Br10\_U2g21 and B1-Br10\_U2g19 we found laccase-like Cu-oxidases encoded that may be involved in the breakdown of phenolic compounds, ring cleavage of aromatic compounds or lignin (Ausec et al., 2011). A near-complete Embden-Meyerhoff pathway was encoded

in all drafts, which include some typical archaeal enzymes such as an ADP-dependent phosphofructokinase/glucokinase (EC 2.7.1.147) and a 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (EC 5.4.2.12; Siebers and Schönheit, 2005). The draft genomes also encoded a “classical” glucose-6-phosphate isomerase as found in most *Eukarya* and *Bacteria*. Moreover, they encoded an eukaryal-type Class I fructose-1,6-biphosphate aldolase (EC 4.1.2.13), which is unusual in *Archaea*, only shared with some members of the *Halobacteria* and *Nanoarchaeota* (Siebers et al., 2001; Say and Fuchs, 2010). The potential for conversion of pyruvate to acetyl CoA was encoded by a pyruvate dehydrogenase complex, while a pyruvate:ferredoxin oxidoreductase (pyruvate synthase; EC 1.2.7.1) was only encoded in the reference genome of “*Candidatus Nanosalinarum*.”

#### *Fermentative lifestyle*

Several genes related to central carbon metabolism and energy production were present in our draft genomes and the four publicly available drafts of *Nanohaloarchaea* described by Ghai et al. (2011), Narasimgarao et al. (2012), and Martínez-García et al. (2014) suggest an anaerobic fermentative lifestyle, which contrasts to the principally aerobic respiratory metabolism suggested previously for “*Candidatus Nanosalina*” and “*Candidatus Nanosalinarum*” by Narasimgarao et al. (2012). First, we found an ADP-forming acetyl-CoA synthase (EC 6.2.1.13), typical for acetate-forming anaerobic fermentative *Archaea* and facultative anaerobic *Halobacteria* (Siebers and Schönheit, 2005) as well as a malic enzyme (EC 1.1.1.38/39), a lactate dehydrogenase (the latter was also reported by Narasimgarao et al., 2012), ferredoxin and a pyruvate:ferredoxin oxidoreductase which suggests at least the potential for incomplete fermentation. Second, while most draft genomes encoded a H<sup>+</sup>-translocating NADH:ubiquinone reductase (EC 1.6.5.3), a plastocyanin-like type I copper protein (most likely involved in electron transfer reactions, e.g., Martín-Cuadrado et al., 2015) and V-type ATPases, none contained identifiable aerobic terminal oxidases. Andrade et al. (2015) most recently reported the absence of cytochrome *c* oxidases in novel *Nanohaloarchaea*-related draft genomes as well. Finally, in contrast to Narasimgarao et al. (2012) we did not find a complete or reduced TCA-cycle in any of the available drafts, only a putative citrate synthase (EC 2.3.3.1) and both subunits of a putative succinyl CoA synthetase (EC 6.2.1.5), as well as a putative citryl CoA ligase (EC 4.1.3.34). In the available genome of “*Candidatus Haloredivivus*” however, the succinyl CoA synthetase was annotated as a citryl-CoA synthetase and had 35% sequence identity (Blast-P; *e*-value 4e-66) to the citryl-CoA synthetase of *Hydrogenobacter thermophilus* (bacterium phylum *Aquificae*). Citryl-CoA synthetase has been shown to act in the latter organism in concert with citryl-CoA lyase as an alternative citrate cleavage pathway to the reductive TCA cycle (Hügler et al., 2007). Hypothetically for the *Nanohaloarchaea*, both enzymes, together with malic enzyme, could perhaps function in the ATP-dependent regeneration of acetyl-CoA and oxaloacetate rather than in the aerobic TCA cycle.

**TABLE 3 | Phylogenetic affiliation and general features of selected reconstructed draft genomes.**

Name <Id>	C	L	GC%	ORF	tRNA	16S	5S	23S	Cont	Str het	Compl
<i>Chitinophagaceae</i> bacterium T5-Br10_B2g13	258	4.30	41	7004	35	0	2	0	2.1	0.0	80.2–91.9
<i>Flavobacteriaceae</i> bacterium T5-Br10_B2g0	132	1.49	33	2494	21	0	0	0	0.9	33.3	61.3–71.2
<i>Rhodothermaceae</i> bacterium Tc-Br11_B2g6_7	109	3.20	62	5220	38	1	1	1	1.1	0.0	79.3–92.8
<i>Halobacteriaceae</i> archaeon Tc-Br11_E2g27	205	2.10	55	3920	25	0	0	0	0.4	0.0	69.3–79.2
<i>Haloferacaceae</i> archaeon PL-Br10_E2g46	171	1.44	51	2586	19	0	0	0	0.5	25.0	67.9–77.4
<i>Haloferacaceae</i> archaeon B1-Br10_E2g22	152	2.37	58	4738	35	0	1	0	3.2	35.7	73.6–83.0
<i>Haloferacaceae</i> archaeon PL-Br10_E2g29	134	1.28	62	2478	21	0	0	0	1.6	44.4	62.3–73.6
<i>Halorubrum</i> sp. PL-Br10_E2g5	160	1.48	65	2756	19	0	0	0	0.4	50.0	79.2–83.0
<i>Haloferacaceae</i> archaeon Tc-Br11_E2g18	139	1.27	67	2372	15	0	0	0	0.0	0.0	62.3–75.5
<i>Nanohaloarchaea</i> archaeon B1-Br10_U2g1	26	0.710	43	1634	29	0	0	1	0.0	0.0	84.9–94.3
<i>Nanohaloarchaea</i> archaeon B1-Br10_U2g21	24	0.816	39	1756	34	0	1	1	2.8	50.0	81.1–86.8
<i>Nanohaloarchaea</i> archaeon B1-Br10_U2g19	35	0.663	41	1472	19	1	0	0	0.0	0.0	81.1–90.6
<i>Nanohaloarchaea</i> archaeon B1-Br10_U2g29	35	0.526	40	1182	19	0	1	0	1.0	0.0	64.2–73.6
<i>Nanohaloarchaea</i> archaeon PL-Br10_U2g5	26	0.84	41	1936	34	1	1	1	0.0	0.0	88.7–94.3
<i>Nanohaloarchaea</i> archaeon PL-Br10_U2g19	55	0.76	42	1702	34	0	2	2	0.9	0.0	81.1–90.6
<i>Nanohaloarchaea</i> archaeon PL-Br10_U2g16	54	0.65	42	1448	18	0	0	0	2.3	20.0	81.1–90.6
<i>Nanohaloarchaea</i> archaeon PL-Br10_U2g27	52	0.58	43	1290	22	3	0	2	0.9	0.0	73.6–77.4
<i>Natrialbaceae</i> archaeon B1-Br10_E2g2	251	2.875	66	5368	30	0	4	0	3.7	11.8	79.2–84.9
<i>Natrialbaceae</i> archaeon B1-Br10_E2g27	207	2.01	60	3710	22	0	1	0	0.8	0.0	62.3–71.7
<i>Natrialbaceae</i> archaeon Tc-Br11_E2g1	210	2.16	65	3990	27	0	0	0	1.8	20.0	75.5–84.9
<i>Natrialbaceae</i> archaeon Tc-Br11_E2g14	136	3.16	63	5916	48	0	0	0	1.6	0.0	84.7–94.3
<i>Natrialbaceae</i> archaeon Tc-Br11_E2g28	216	2.33	61	4342	34	0	2	1	0.3	0.0	71.7–86.8
<i>Natrialbaceae</i> archaeon Tc-Br11_E2g8	192	2.29	69	4206	32	0	1	0	0.7	16.7	71.7–83.0
<i>Natrialbaceae</i> archaeon PL-Br10_E2g26	167	1.46	61	2646	18	0	2	0	0.0	0.0	67.9–81.1

C, # contigs; L, length in Mb; GC%, the average GC content; ORF, # open reading frames (CDS); tRNA, # tRNA sequences; 16S, 5S, and 23S = resp. # rRNA; Cont and Str het = % of bin contamination and strain heterogeneity estimated with CheckM; Compl = lower and upper bounds of predictions of % completeness based on 53 archaeal core genes or 111 single copy genes for the bacterial drafts.

## Halobacteria

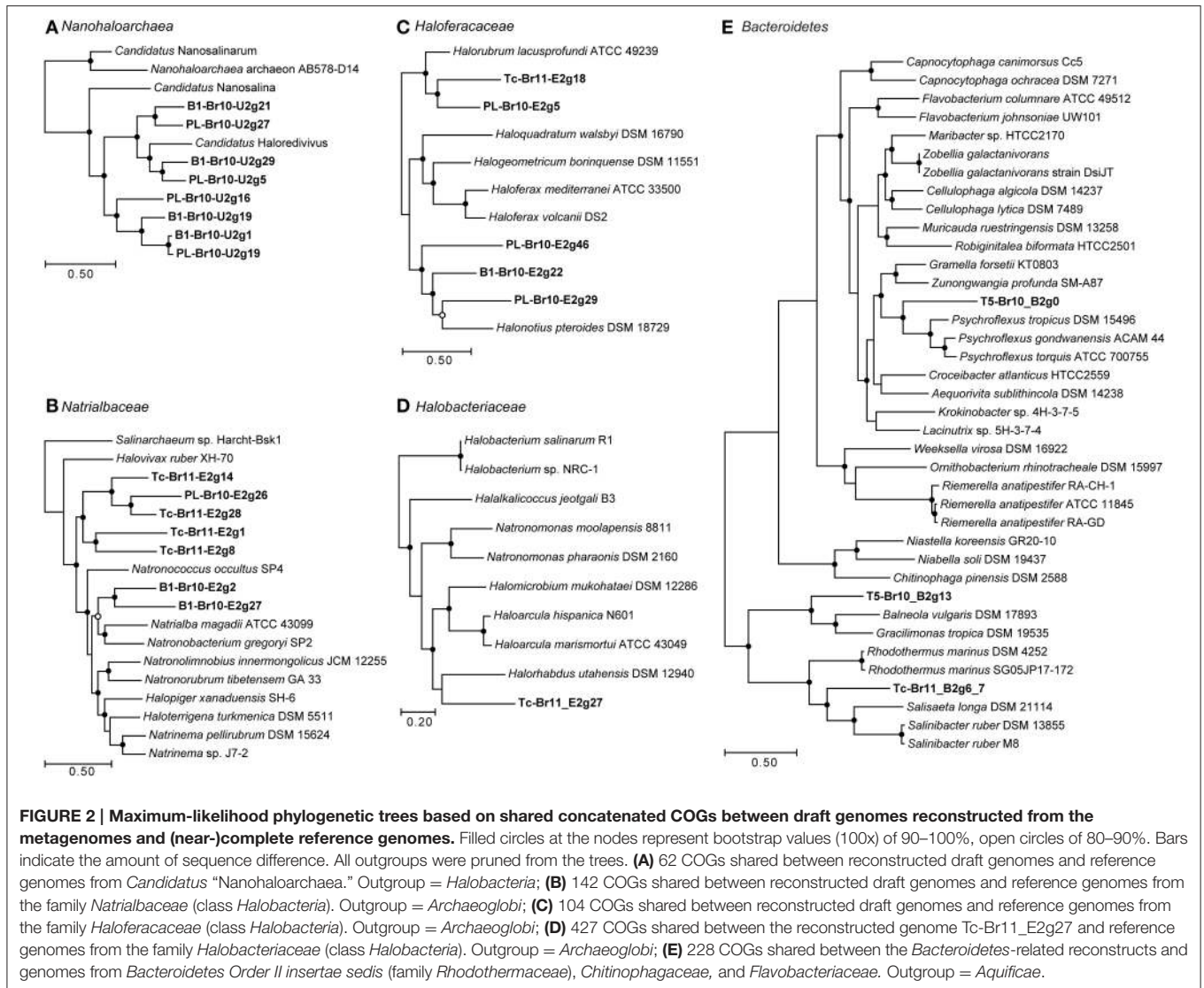
### Phylogeny

Thirteen draft genomes from members of all three known orders within the class *Halobacteria* (as recently defined in Gupta et al., 2014) were obtained from the datasets B1-Br10, Tc-Br11 and PL-Br10 of the most saline soda brines (Table 3). Within the family *Natrialbaceae*, the two drafts B1-Br10\_E2g2&27 were affiliated with the genera *Natronococcus*, *Natrialba*, and *Natronobacterium* (Figure 2B; Supplementary Table 5). Since, 16S rRNA sequences belonging to these genera were found in the corresponding amplicon datasets of B1, but their ANI and the ANIs with the three complete reference genomes shown in our genome tree were only between 74.3 and 77.4%, it is not clear whether the drafts belong to one of these three genera or represent one or two novel, distinct genera. Five more drafts of *Natrialbaceae*-related organisms were reconstructed that, based on our genome tree, appeared to be more related to each other than to any other complete genome sequenced from this family. The drafts Tc-Br11\_E2g28 and PL-Br10\_E2g26 belonged to the same species (ANI = 98.6%) thriving in different brines. Within the family *Haloferacaceae*, the two drafts Tc-Br11\_E2g18 and PL-Br10\_E2g5 were affiliated with the genus *Halorubrum* (Figure 2C; Supplementary Table 6). Since the latter had an ANI of 80.6% with *Halorubrum lacusprofundi* ATCC 49239, it almost

certainly belongs to the genus *Halorubrum*. The three drafts PL-Br10\_E2g29&46 and B1-Br10\_E2g22 were affiliated with the genus *Halonotius*, but most likely belonged to three distinct, novel genera (ANIs between 69.9 and 74.6%). One draft genome obtained from the trona crystallizer Tc, Tc-Br11\_E2g27 belonged to the family *Halobacteriaceae* (Figure 2D; Supplementary Table 7) and comprised a distinct genus related to *Halorhabdus* (ANI = 68.7%).

### Rhodopsins and metabolic versatility

All reconstructed draft genomes encoded for versatile metabolisms indicating aerobic heterotrophy. Almost all draft genomes possessed a halorhodopsin and/or light-activated signal transducers (sensory rhodopsin I and II) typical of extremely halophilic *Archaea* (Supplementary Table 9). Most drafts encoded a near-complete Krebs cycle and the potential conversion of pyruvate to acetyl CoA by a pyruvate synthase or in some cases by a pyruvate dehydrogenase complex. All drafts encoded subunits I and/or III of a cytochrome *c* oxidase most likely functioning as a terminal oxidase for aerobic respiration (Supplementary Table 10). Almost all genomes encoded a putative L-lactate dehydrogenase (EC 1.1.1.27), which might point to the ability for sugar fermentation. Notably, all of the *Natrialbaceae*-related drafts as well as three of the



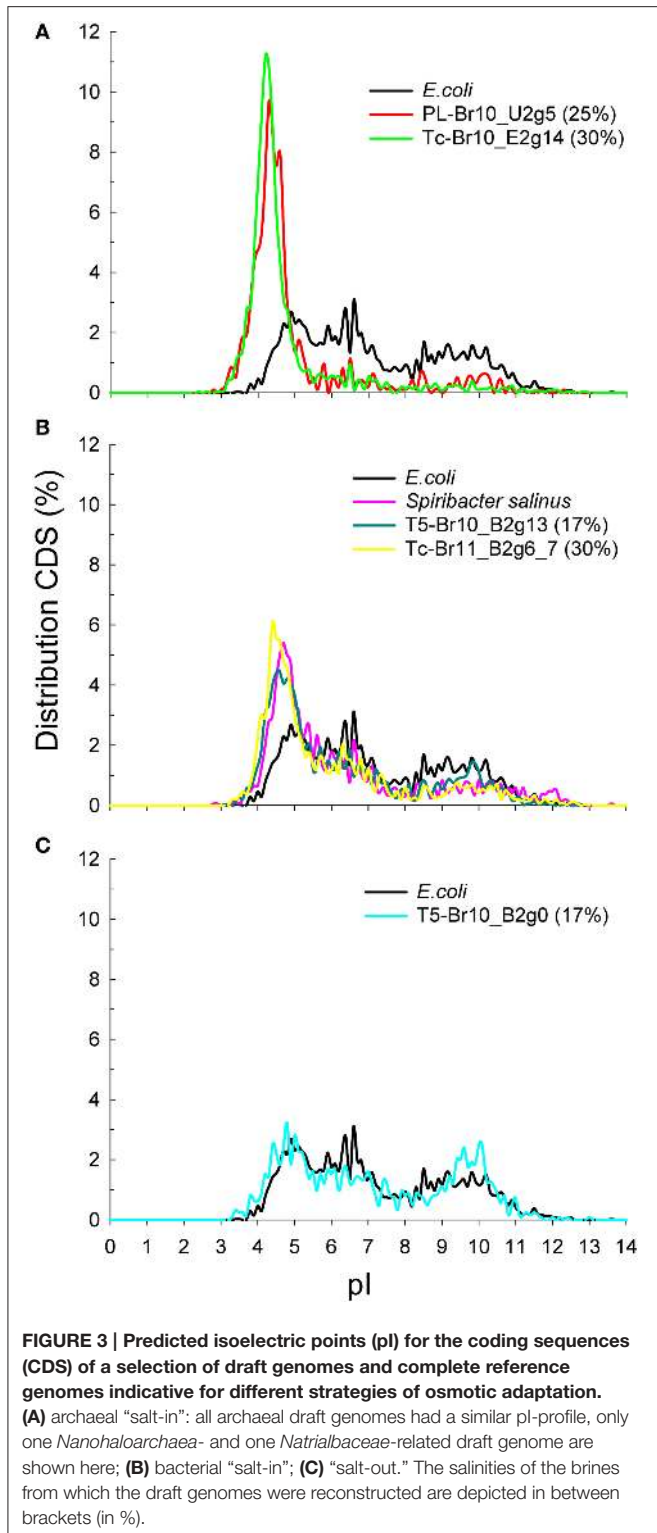
**FIGURE 2 | Maximum-likelihood phylogenetic trees based on shared concatenated COGs between draft genomes reconstructed from the metagenomes and (near)-complete reference genomes.** Filled circles at the nodes represent bootstrap values (100x) of 90–100%, open circles of 80–90%. Bars indicate the amount of sequence difference. All outgroups were pruned from the trees. **(A)** 62 COGs shared between reconstructed draft genomes and reference genomes from *Candidatus* “Nanohaloarchaea.” Outgroup = *Halobacteria*; **(B)** 142 COGs shared between reconstructed draft genomes and reference genomes from the family *Natrialbaeaceae* (class *Halobacteria*). Outgroup = *Archaeoglobi*; **(C)** 104 COGs shared between reconstructed draft genomes and reference genomes from the family *Haloferacaceae* (class *Halobacteria*). Outgroup = *Archaeoglobi*; **(D)** 427 COGs shared between the reconstructed genome Tc-Br11\_E2g27 and reference genomes from the family *Halobacteriaceae* (class *Halobacteria*). Outgroup = *Archaeoglobi*; **(E)** 228 COGs shared between the *Bacteroidetes*-related reconstructs and genomes from *Bacteroidetes* Order *II insertae sedis* (family *Rhodothermaceae*), *Chitinophagaceae*, and *Flavobacteriaceae*. Outgroup = *Aquificae*.

*Haloferacaceae*-related drafts encoded a tungsten-containing aldehyde:ferredoxin oxidoreductase (AOR; EC 1.2.7.5), a widespread enzyme amongst facultative fermenters (Hagedoorn et al., 2005). Five of the halobacterial drafts encoded the catalytic subunit PsrA of a putative molybdenum-containing thiosulfate/polysulfide reductase (PsrABC; Supplementary Table 10), which can catalyze the quinone-coupled reduction of polysulfides ( $S_n^{2-}$ ) in the bacterium *Thermus thermophilus* (class *Deinococci*) thriving in geothermal hot springs (Jormakka et al., 2008). Furthermore, membrane-bound polysulfide reductases were expressed during growth of *Halanaeroarchaeum sulfurireducens* HSR2, a sulfidogenic, extreme halophilic, neutrophilic, obligate anaerobic member of the *Halobacteriaceae* capable of dissimilatory reduction of elemental sulfur using acetate as the sole electron donor (Sorokin et al., 2015c). Polysulfides are unstable intermediates formed by the chemical reaction of elemental sulfur with free sulfide in the pH-neutral environment from which *H. sulfurireducens* was obtained. In

contrast, thiosulfate and polysulfides are the expected abundantly present sulfur intermediates in alkaline environments such as soda brines. Four of the five PsrABC-containing drafts also encoded rhodanese-like sulfurtransferases (EC:2.8.1.1), which are suggested to be involved in the binding, stabilization and transfer of polysulfides to polysulfide reductase in *H. sulfurireducens* HSR2 (Sorokin et al., 2015c). Additionally, *Haloferacaceae* archaeon Tc-Br11\_E2g18 encoded the electron-transfer subunit (NarH) of a putative membrane-bound nitrate reductase (Nar). The molybdoenzyme Nar has been previously shown to function as a terminal electron acceptor in *Haloarcula marismortui* and *Haloferax* sp. (families *Halobacteriaceae* and *Haloferacaceae* respectively) under denitrifying conditions (reviewed by Cabello et al., 2004).

### Central carbon metabolism

The glycolytic pathway within their central carbon metabolisms appeared to be rather versatile as well. From all three



families draft genomes were obtained that encoded a 2-keto-3-deoxygluconate kinase (KDG kinase; EC 2.7.1.45), the key-enzyme of non- or semi-phosphorylative Entner-Doudoroff pathway for glucose degradation (Siebers and Schönheit, 2005), although identifiable glucose dehydrogenases

and gluconate dehydratases (enzymes that lead to the formation of 2-keto-3-deoxygluconate) were lacking in our incomplete drafts. Interestingly, most drafts contained a putative glucose/sorbosone dehydrogenase (COG2133), but this enzyme has to our knowledge not been characterized in a member of the *Halobacteria*. Three of the *Natrialbaeae*-related drafts contained within the *Natrialbaeae* encoded a putative non-phosphorylative glyceraldehyde 3-phosphate dehydrogenase (gapN; EC 1.2.1.9), the other ten drafts contained one or both types of phosphorylative glyceraldehyde 3-phosphate dehydrogenase(s), namely GAPDH (EC 1.2.1.12) and/or gap2 (EC 1.2.1.59). The latter enzyme is part of the Emden-Meyerhoff pathway for hexose degradation, which has been shown in *Halococcus saccharolyticus* (family *Halobacteriaceae*) to be the main pathway for fructose degradation (Johnsen et al., 2001). The presence of a similarly branched Entner-Doudoroff/Emden-Meyerhoff pathway is suspected to be the general case, since all drafts contained at least one of the enzymes typical for a modified Emden-Meyerhoff pathway, which includes an ATP-dependent glucokinase, an ADP-dependent glucokinase/phosphofructokinase and a fructose-bisphosphate aldolase (both Class I and II found).

#### Primary organic carbon degradation

Polysaccharide-degrading potential was found in almost all *Halobacteria*-related draft genomes (Supplementary Table 11). Almost all drafts encoded glucoamylases and related glycoside hydrolases. Putative chitinases were encoded by *Natrialbaeae* archaeon B1-Br10\_E2g27 and *Haloferacaceae* archaeon Tc-Br11\_E2g18, with catalytic domains resembling those of GH18 and GH19 CAZy family proteins respectively. Putative cellulases were encoded by the *Natrialbaeae* archaea B1-Br10\_E2g27, Tc-Br11\_E2g14 and Tc-Br11\_E2g28, and *Haloferacaceae* archaeon PL-Br10\_E2g29. These particular organisms might be able to hydrolyze or even grow on recalcitrant organic substrates such as cellulose and chitin as shown previously for several *Halobacteria* isolated from soda lakes (Sorokin et al., 2015d). Some of these isolates remarkably grew both on celluloses and chitin. Interestingly, also draft genome B1-Br10\_E2g27 held both cellulolytic and chitinolytic potential.

#### Aerobic CO oxidation

The draft genome of *Natrialbaeae* archaeon Tc-Br11\_E2g1 contained *coxM* and *coxS* genes, the draft of *Natrialbaeae* archaeon Tc-Br11\_E2g28 contained a *coxL*. Together these genes encode the three main subunits of a putative aerobic carbon monoxide dehydrogenase (CODH). The large subunit (*coxL*) of CODH is a molybdopterin dehydrogenase FAD-binding protein with a Mo-Cu sulfur-bridged cluster in its active site that catalyzes the oxidation of CO to CO<sub>2</sub>, yielding two electrons and two protons (Dobbek et al., 2001; Gnida et al., 2003; King and Weber, 2007). While the use of CO as a sole energy and carbon source or as an alternate energy source coupled to CO<sub>2</sub> fixation ("carboxydrotrophy") is typically restricted to few bacterial groups and can occur only under elevated CO concentrations, the use of CO at low concentrations as an alternative energy source ("carboxydovory") is suggested to be widespread amongst

heterotrophs (King and Weber, 2007; Cunliffe, 2011; King, 2013). It was recently proposed that aerobic CO oxidation might have evolved in the phylum *Euryarchaeota* (genus *Natronorubrum*; King, 2013) and CO consumption has been demonstrated in isolates from the genus *Halorubrum* and *Natronorubrum* (King, 2015). Since, gammaproteobacterial carboxydrotrophs as well as carboxydovores have been isolated from several soda lakes around the world (Sorokin et al., 2015a), CO must be to some extent available to soda lake aerobic microbes. As suggested previously for heterotrophic bacteria in the oceans (Martin-Cuadrado et al., 2009), *Halobacteria* might energetically benefit as well from metabolizing CO released by the photolysis of organic material in the surface waters of soda lakes. However, whether any members of the *Euryarchaeota*, including members of the *Halobacteria*, can actually use CO as an energy source remains to be proven.

## **Bacteroidetes**

### **Phylogeny**

Three *Bacteroidetes*-related draft genomes were obtained from Tc-Br11 and PL-Br10 (Table 3; Figure 2E; Supplementary Table 8). The draft genome T5-Br10\_B2g13 was affiliated with the genera *Gracilimonas* and *Balneola* (order Sphingobacteriales; Chitinophagaceae, no rank; Urios et al., 2006; Choi et al., 2009), is only distantly related to the orders *Flavobacteriales* and *Bacteroidetes Order II Incertae sedis* and had an average GC content of 41%. Similar phylogeny and GC content was reported for *Gracilimonas* and the related genus *Balneola* and annotated genes were affiliated consistently with these two genera (best-hit analysis). While *Gracilimonas*-related OTUs comprised up to 18.84% of total 16S rRNA amplicon sequences from T5, 2.50% of *Chitinophagaceae*-related OTUs could not be assigned to a known genus (Table 2). Since, the ANIs of T5-Br10\_B2g13 with the incomplete genomes available from *Balneola vulgaris* DSM 17893 and *Gracilimonas tropica* DSM 19535 were low (65.2 and 66.2% respectively), T5-Br10\_B2g13 must belong to a novel genus within the *Chitinophagaceae*. Its closest cultured relative appears to be a proteolytic satellite (bacterial symbiont) of a soda lake filamentous cyanobacterium isolated from the same area (DY Sorokin, unpublished data). Notably in our genome tree, *Gracilimonas*, *Balneola*, and T5-Br10\_B2g13 formed a separate clade from the other three genera within the family *Chitinophagaceae* (namely *Chitinophaga*, *Niabella*, and *Niastella*).

The draft genome T5-Br10\_B2g0 was affiliated with the genus *Psychroflexus* (family *Flavobacteriaceae*; Bowman et al., 1998; Donachie et al., 2004; Yoon et al., 2009) in our genome tree. Since no *Psychroflexus*-related OTUs were found in the *Flavobacteriaceae*-related 16S rRNA amplicon sequences of T5 (Table 2) and ANIs of T5-Br10\_B2g0 with the incomplete genomes available from *Psychroflexus gondwanensis* ACAM 44 and *Psychroflexus tropicus* DSM 15496 were low (68.4 and 68.2% respectively), T5-Br10\_B2g0 must represent the draft genome of a novel lineage within the *Flavobacteriaceae*.

The draft genome Tc-Br11-B2-g6\_7 belonged to the family *Rhodothermaceae* (*Bacteroidetes Order II Incertae sedis*, currently order *Cytophagales*, class *Cytophagia*; Park et al., 2014) branching in a genome tree between the genera *Rhodothermus* and *Salisaeta*

of which halophilic isolates have been obtained from marine hot springs and high salt environments respectively (Antón et al., 2002; Vaisman and Oren, 2009). *Rhodothermaceae*-related OTUs without genus assignment as well as *Salisaeta*-related OTUs were found in the 16S rRNA amplicon sequences from Tc (0.24 and 0.06% of total OTUs respectively, Table 2). Since the ANI of Tc-Br-Bacteroidetes-g6\_7 with the incomplete genome available from *Salisaeta longa* DSM 21114 was low (67.4%), Tc-Br11-B2-g6\_7 belongs to a novel extreme halophilic *Bacteroidetes* lineage within the *Rhodothermaceae*.

### **Central carbon metabolism and primary organic carbon degradation**

In all *Bacteroidetes* draft genomes cytochrome c oxidase subunits I and III, marker genes for aerobic respiration, were encoded (Supplementary Table 12), as well as a complete or a near-complete Embden-Meyerhof pathway, pyruvate dehydrogenase complex, and Krebs cycle. Genes encoding both subunits of the 2-oxoglutarate:ferredoxin oxidoreductase (EC 1.2.7.3) as well as a gene encoding a L-lactate dehydrogenase were found for *Chitinophagaceae* bacterium T5-Br10\_B2g13 and *Rhodothermaceae* bacterium Tc-Br11\_B2g6\_7, suggesting they have the potential for sugar fermentation under anoxic or micro-oxic conditions.

Several putative glycoside hydrolases involved in polysaccharide degradation were encoded in the three *Bacteroidetes* draft genomes (Supplementary Table 13). Putative glucoamylases, cellulases, and related enzymes suggest amyolytic and cellulolytic potential for all three organisms. Hemicellulases and related enzymes were only found in *Chitinophagaceae* bacterium T5-Br10\_B2g13 and *Rhodothermaceae* bacterium Tc-Br11\_B2g6\_7. Both *Chitinophagaceae* bacterium T5-Br10\_B2g13 and *Flavobacteriaceae* bacterium T5-Br10\_B2g0 possessed a putative  $\beta$ -galactosidase (*lacZ*) and functionally related glycoside hydrolases, as well as chitinase related enzymes (CAZy family GH 23).

### **Denitrification: N<sub>2</sub>O reduction**

The genome of *Chitinophagaceae* bacterium T5-Br10\_B2g13 encodes a key functional gene *nosZ* for dissimilatory nitrous oxide (N<sub>2</sub>O) reduction within the denitrification pathway (Wood et al., 2001; Jones et al., 2013; Supplementary Table 12). According to the model of *Geobacillus thermodenitrificans*, the gene also lacked a characteristic Tat-signal at the N-terminus for translocation across the membrane (Liu et al., 2008). This finding agrees with the known widespread N<sub>2</sub>O reduction potential in various environments by members of the *Firmicutes* and *Bacteroidetes* which is encoded by a *nosZ* with a presumed Sec-translocation pathway (Jones et al., 2013). Although it is known that some halophilic *Halobacteria* have the capacity to reduce N<sub>2</sub>O (Jones et al., 2013), the process was previously only shown for soda lake isolates within the *Gammaproteobacteria* belonging to the genera *Halomonas*, *Thioalkalivibrio*, and *Alkalispirillum* (Sorokin et al., 2012, 2014). Amongst the >5 kb contigs we found *nosZ* genes in the three most saline datasets related to the genera *Halorubrum* and *Natronococcus*, but we did not find gammaproteobacterial *nosZ* genes in the >5 kb contigs of T5.

### Osmotic adaptation: rhodopsins, salt-in strategy and organic osmolytes

Finally, both *Chitinophagaceae* bacterium T5-Br10\_B2g13 and *Flavobacteriaceae* bacterium T5-Br10\_B2g0 contained a putative sodium pumping rhodopsin with characteristic NQ motif, which could aid in keeping the intracellular sodium ion concentration low (Inoue et al., 2013; Kwon et al., 2013; Supplementary Table 9). *Flavobacteriaceae* bacterium T5-Br10\_B2g0 possesses a putative proton pumping proteorhodopsin as well (Béja et al., 2001). Maintaining osmotic balance with high levels of cytoplasmic KCl, often referred to as a “salt-in strategy” (Oren, 2013), appears to be a common strategy for members of the *Halobacteria* and *Nanohaloarchaea* thriving in hypersaline habitats (Ghai et al., 2011; Youssef et al., 2014; **Figure 3**; Supplementary Figure 3). *Chitinophagaceae* bacterium T5-Br10\_B2g13 and *Rhodothermaceae* bacterium Tc-Br11\_B2g6\_7 seem to deploy this strategy to some extent since their predicted proteomes were enriched in acidic residues compare to that of *E. coli*, despite the fact that putative genes for the synthesis and uptake of organic osmolytes (including betaine aldehyde dehydrogenases and betaine/carnitine/choline transporters) were found in their draft genomes. So far, the “salt-in strategy” was shown only for two bacterial phyla, in the orders *Halanaerobiales* and *Natranaerobiales* (*Firmicutes*; Oren, 2013) and *Salinibacter* (*Bacteroidetes*; Mongodin et al., 2005). No acid-shifted proteome compared to *E. coli* was found for *Flavobacteriaceae* bacterium T5-Br10\_B2g0.

### Natural Abundance of the Reconstructed Genomes

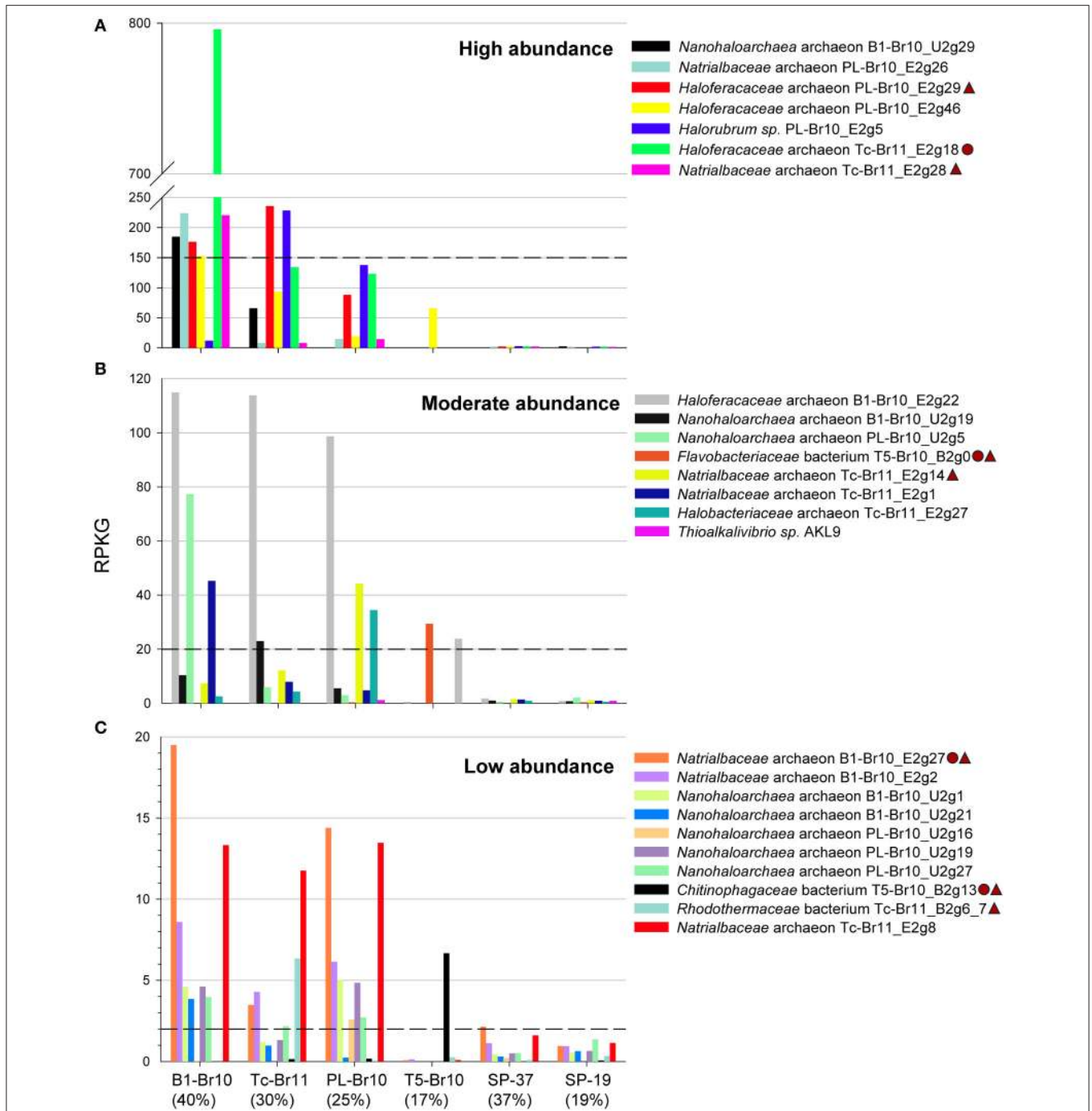
In agreement with the 16S amplicon sequencing, *Haloferacaceae*-related genomes were the most abundant in the three most saline brines of B1, Tc and PL (**Figure 4**). Although the putative chitinolytic *Haloferacaceae archaeon* Tc-Br11\_E2g18 was by far the most abundant organism in the brine of B1, the most saline brine still harbored a considerable diverse population of other *Halobacteria* (**Figure 4A**). The dominant high-GC *Halobacteria* in the most saline brines left their signature in the GC profiles of the raw metagenomic reads (Supplementary Figure 2). *Nanohaloarchaea*-related genomes were confined to the most saline brines and were moderately (reads per kilobase of sequenced reads per gigabase of mapped reads (RPKG) between 120 and 20, **Figure 4B**) or low abundant (RPKG between 20 and 2, **Figure 4C**). Previous nanohaloarchaeal metagenomic genome reconstructions were only obtained from more moderate salinities below 250 g/L of total salts (Ghai et al., 2011; Narasingarao et al., 2012; Martínez-García et al., 2014). Only one of the *Halobacteria*-related genomes, *Haloferacaceae archaeon* PL-Br10\_E2g46, was moderately abundant in the brine of T5 with a salinity of 170 g/L. It was even more abundant than the *Bacteroidetes*-related genomes that were confined to the brine of T5 (using an arbitrary cut-off of RPKG=2 for presence of a certain organism within a dataset) and *Thioalkalivibrio* sp. AKL9 (class *Gammaproteobacteria*). *Rhodothermaceae* bacterium Tc-Br11\_B2g6\_7 had only a low abundance in the brine of Tc. The draft genomes that were reconstructed from a different brine, but were identified to be the same species, *Natrialbaeae archaeon* PL-Br10\_E2g26 &

Tc-Br11\_E2g28 and *Nanohaloarchaea archaeon* B1-Br10\_U2g1 and PL-Br10\_U2g19, showed comparable RPKG for all four datasets. Finally, only very low read abundance of our soda lake draft genomes was found in the metagenomics datasets SP-37 and SP-19, originating from salterns with neutral pH. Although *Halobacteria*, *Nanohaloarchaea*, and *Bacteroidetes* are common taxa in high-salt environments, the specific organisms belonging to these groups that can thrive at high alkalinities are likely to differ from those that thrive at circum-neutral pH. The same holds true for pure cultures, where the isolates from soda lakes in most cases differentiate from their halophilic counterparts on the genus level (Sorokin et al., 2014, 2015a).

## CONCLUSIONS AND FUTURE RESEARCH

Salinity has a similar dominant influence on the overall prokaryote community structure of alkaline soda lake brines as it has on that of hypersaline ponds of solar salterns with neutral pH; over a salt concentration of roughly 250 g/L bacterial abundance declines and dominance shifts to members of the *Halobacteria* (Rodríguez-Valera et al., 1985; Oren, 1994; Ghai et al., 2011). Yet hypersaline soda lake brines with salinities exceeding 250 g/L harbor considerably different and more diverse communities compared to solar saltern brines of similar salinities. The absence of calcium salts that precipitate at high pH and much lower prevalent magnesium concentrations could offer explanations for the differences found in the community structure. Furthermore, the investigated, shallow soda lakes from the Kulunda Steppe have a very unstable water regime with large fluctuations in salinity throughout the seasons and the year, according to local snow melt, groundwater inflow and evaporation rates. These fluctuations could also contribute to sustaining such diverse soda brine communities. Finally, the nearly 50% lower osmotic pressure in hypersaline soda lake brines compared to solar salterns (Banciu and Sorokin, 2013; Sorokin et al., 2015a) might be the reason why in particular *Nanohaloarchaea*, which rely solely on the salt-in strategy for osmotic adaption, prefer higher salinities in alkaline brines than in NaCl brines with circum-neutral pH.

By reconstructing the first draft genomes from members of the candidate class *Nanohaloarchaea* present in hypersaline soda brines, we extend the amount of available genomic data from this enigmatic new group of *Archaea* considerably and suggest that their inferred ecological lifestyle with respect to oxygen respiration should be revised. Furthermore, we identified hydrolytic capacity not only within previously unrecognized members of the *Bacteroidetes*, but also amongst members of the *Halobacteria* which outnumber the first in the most saline brines. The fact that some members of the *Halobacteria* might fulfill a primary rather than a secondary degrading function in the organic carbon cycle of soda lakes is supported by previous culture-dependent studies. Overall, the dominant organisms in the most saline brines seem to be aerobic heterotrophs that adapted to the presence of light and to the low solubility of oxygen at high salinities with putative sugar fermentative capabilities. We identified two members of the *Bacteroidetes* that might utilize light not simply as an additional energy source or



**FIGURE 4 |** Read recruitments (<95% N-identity) of reconstructed draft genomes and a reference draft genome from the genus *Thioalkalivibrio* from different metagenomic datasets. SP-37 and SP-19 are metagenomes from a crystallizer and an intermediate saline pond of the solar saltern in Santa Pola, Spain, with 37 and 19% salinity respectively (Ghai et al., 2011). The salinity (in %) for the different soda brine metagenomes is given between brackets. Reconstructed draft genomes encoding putative chitinases are marked in the figure legend with a circle, those encoding putative cellulases with a triangle. RPKG, Reads Per Kilobase of sequenced reads per Gigabase of mapped reads. **(A)** Putative organisms with relatively high abundance (> 150 RPKG), **(B)** with moderate abundance, **(C)** and with relatively low abundance (< 20 RPKG).

sensory input for movement, but rather to pump out Na<sup>+</sup> via sodium pumping rhodopsins.

The presented genomic data provide a first insight into the physiological potential of a wealth of still uncultured

organisms living in hypersaline soda lake brines. We hope this study can pave the way for the design of novel isolation strategies and protein structure-function studies, crucial in obtaining unequivocal evidence for the proposed metabolic

features of the novel organisms. More in depth sampling over time or space of soda brines could bring a more detailed understanding of the influence of environmental gradients in salinity and salt composition, as well as of other parameters such as pH and seasonality, on specific microbial populations and functional groups. Future research should combine transcriptomic and/or proteomic approaches with *in situ* rates of element transformations to reveal the active prokaryote community members in hypersaline soda lake brines in relation to salinity and could aim at verifying the functional role and relative importance of rare vs. abundant community members.

## AUTHOR CONTRIBUTIONS

DS, PH, and GM were responsible for the conception of the work. DS collected the samples and extracted the DNA. GM prepared samples for amplicon sequencing and performed subsequent analysis. ST was responsible for the DNA sequencing. RG performed the assembly of the metagenomic reads and assisted CV with the genome reconstructions and subsequent analysis. CV integrated and interpreted all data and is the primary author of the manuscript. She received thereby substantial contributions

of all co-authors, including final approval for this version to be published.

## FUNDING

CV and GM are supported by the ERC Advanced Grant PARASOL (No. 322551). The work of CV was further supported by an STSM Grant from the COST Action ES1103. RG is partially supported by the Grant Agency of the Czech Science Foundation under the research grant 13-00243S. FR-V and RG are supported by a grant from the European Union (MaCuMBA project, grant 311975). DS is supported by the Russian Foundation for Basic Research (16-04-00035). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmich.2016.00211>

## REFERENCES

- Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31, 533–538. doi: 10.1038/nbt.2579
- Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146. doi: 10.1038/nmeth.3103
- Andrade, K., Logemann, J., Heidelberg, K. B., Emerson, J. B., Comolli, L. R., Hug, L. A., et al. (2015). Metagenomic and lipid analyses reveal a diel cycle in a hypersaline microbial ecosystem. *ISME J.* 9, 2697–2711. doi: 10.1038/ismej.2015.66
- Antón, J., Oren, A., Benlloch, S., Rodríguez-Valera, F., Amann, R., and Rosselló-Mora, R. (2002). *Salinibacter ruber* gen. nov., sp. nov., a novel, extremely halophilic member of the bacteria from saltern crystallizer ponds. *Int. J. Syst. Evol. Microbiol.* 52, 485–491. doi: 10.1099/00207713-52-2-485
- Antón, J., Rosselló-Mora, R., Rodríguez-Valera, F., and Amann, R. (2000). Extremely halophilic bacteria in crystallizer ponds from solar salterns. *Appl. Environ. Microbiol.* 66, 3052–3057. doi: 10.1128/AEM.66.7.3052-3057.2000
- Asao, M., Pinkart, H. C., and Madigan, M. T. (2011). Diversity of extremophilic purple phototrophic bacteria in soap lake, a central washington (USA) soda lake. *Environ. Microbiol.* 13, 2146–2157. doi: 10.1111/j.1462-2920.2011.02449.x
- Ausec, L., Zakrzewski, M., Goesmann, A., Schlüter, A., and Mandic-Mulec, I. (2011). Bioinformatic analysis reveals high diversity of bacterial genes for laccase-like enzymes. *PLoS ONE* 6:e25724. doi: 10.1371/journal.pone.0025724
- Banciu, H. L., and Sorokin, D. Y. (2013). “Adaptation in Haloalkaliphiles and Natronophilic bacteria,” in *Polyextremophiles*, eds J. Seckbach, A. Oren, and H. Stan-Lotter (Dordrecht: Springer), 121–178.
- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., et al. (2004). The Pfam protein families database. *Nucleic Acids Res.* 32(Suppl. 1), D138–D141. doi: 10.1093/nar/gkh121
- Bauer, M., Kube, M., Teeling, H., Richter, M., Lombardot, T., Allers, E., et al. (2006). Whole genome analysis of the marine *Bacteroidetes* ‘*Gramella forsetii*’ reveals adaptations to degradation of polymeric organic matter. *Environ. Microbiol.* 8, 2201–2213. doi: 10.1111/j.1462-2920.2006.01152.x
- Béja, O., Spudich, E. N., Spudich, J. L., Leclerc, M., and DeLong, E. F. (2001). Proteorhodopsin phototrophy in the ocean. *Nature* 411, 786–789. doi: 10.1038/35081051
- Benlloch, S., López-López, A., Casamayor, E. O., Øvreås, L., Goddard, V., Daae, F. L., et al. (2002). Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. *Environ. Microbiol.* 4, 349–360. doi: 10.1046/j.1462-2920.2002.00306.x
- Boldareva, E. N., Akimov, V. N., Boychenko, V. A., Stadnichuk, I. N., Moskalenko, A. A., Makhneva, Z. K., et al. (2008). *Rhodobaca barguzinensis* sp. nov., a new alkaliphilic purple nonsulfur bacterium isolated from a soda lake of the Barguzin Valley (Buryat Republic, Eastern Siberia). *Microbiology* 77, 206–218. doi: 10.1134/S0026261708020148
- Boldareva, E. N., Bryantseva, I. A., Tsapin, A., Nelson, K., Sorokin, D. Y., Tourova, T. P., et al. (2007). The new alkaliphilic bacteriochlorophyll a-containing bacterium *Roseinatronobacter monicus* sp. nov. from the hypersaline Soda Mono Lake (California, United States). *Microbiology* 76, 82–92. doi: 10.1134/S0026261707010122
- Bowman, J. P., McCammon, S. A., Lewis, T., Skerratt, J. H., Brown, J. L., Nichols, D. S., et al. (1998). *Psychroflexus torquis* gen. nov., sp. nov. a psychrophilic species from Antarctic sea ice, and reclassification of *Flavobacterium gondwanense* (Dobson et al., 1993) as *Psychroflexus gondwanense* gen. nov., comb. nov. *Microbiology* 144, 1601–1609. doi: 10.1099/00221287-144-6-1601
- Cabello, P., Roldán, M. D., and Moreno-Vivián, C. (2004). Nitrate reduction and the nitrogen cycle in *Archaea*. *Microbiology* 150, 3527–3546. doi: 10.1099/mic.0.27303-0
- Casamayor, E. O., Calderón-Paz, J. I., and Pedrós-Alió, C. (2000). 5S rRNA fingerprints of marine *Bacteria*, halophilic *Archaea* and natural prokaryotic assemblages along a salinity gradient. *FEMS Microbiol. Ecol.* 34, 113–119. doi: 10.1111/j.1574-6941.2000.tb00760.x
- Casamayor, E. O., Massana, R., Benlloch, S., Øvreås, L., Díez, B., Goddard, V. J., et al. (2002). Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. *Environ. Microbiol.* 4, 338–348. doi: 10.1046/j.1462-2920.2002.00297.x
- Castelle, C. J., Wrighton, K. C., Thomas, B. C., Hug, L. A., Brown, C. T., Wilkins, M. J., et al. (2015). Genomic expansion of domain archaea highlights roles for



- organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* 25, 690–701. doi: 10.1016/j.cub.2015.01.014
- Cho, Y., Chung, H., Jang, G. I., Choi, D. H., Noh, J. H., and Cho, B. C. (2013). *Gracilimonas rosea* sp. nov., isolated from tropical seawater, and emended description of the genus *Gracilimonas*. *Int. J. Syst. Evol. Microbiol.* 63, 4006–4011. doi: 10.1099/ijs.0.052340-0
- Choi, D. H., Zhang, G. I., Noh, J. H., Kim, W. S., and Cho, B. C. (2009). *Gracilimonas tropica* gen. nov., sp. nov., isolated from a Synechococcus culture. *Int. J. Syst. Evol. Microbiol.* 59, 1167–1172. doi: 10.1099/ijs.0.005512-0
- Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 41(Database issue):gkt1244. doi: 10.1093/nar/gkt1244
- Cox, M. P., Peterson, D. A., and Biggs, P. J. (2010). SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11:485. doi: 10.1186/1471-2105-11-485
- Cunliffe, M. (2011). Correlating carbon monoxide oxidation with *cox* genes in the abundant marine Roseobacter clade. *ISME J.* 5, 685–691. doi: 10.1038/ismej.2010.170
- De Castro, R. E., Maupin-Furlow, J. A., Giménez, M. I., Seitz, M. K. H., and Sánchez, J. J. (2006). Haloarchaeal proteases and proteolytic systems. *FEMS Microbiol. Rev.* 30, 17–35. doi: 10.1111/j.1574-6976.2005.00003.x
- Dillon, J. G., Carlin, M., Gutierrez, A., Nguyen, V., and McLain, N. (2013). Patterns of microbial diversity along a salinity gradient in the Guerrero Negro solar saltern, Baja CA Sur, Mexico. *Front. Microbiol.* 4:399. doi: 10.3389/fmicb.2013.00399
- Dimitriu, P. A., Pinkart, H. C., Peyton, B. M., and Mormile, M. R. (2008). Spatial and temporal patterns in the microbial diversity of a meromictic soda lake in Washington state. *Appl. Environ. Microbiol.* 74, 4877–4888. doi: 10.1128/AEM.00455-08
- Dobbek, H., Svetlichnyi, V., Gremer, L., Huber, R., and Meyer, O. (2001). Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] cluster. *Science* 293, 1281–1285. doi: 10.1126/science.1061500
- Donachie, S. P., Bowman, J. P., and Alam, M. (2004). *Psychroflexus tropicus* sp. nov., an obligately halophilic Cytophaga-Flavobacterium-Bacteroides group bacterium from an Hawaiian hypersaline lake. *Int. J. Syst. Evol. Microbiol.* 54, 935–940. doi: 10.1099/ijs.0.02733-0
- Fernández, A. B., Ghai, R., Martín-Cuadrado, A., Sánchez-Porro, C., Rodríguez-Valera, F., and Ventosa, A. (2014a). Prokaryotic taxonomic and metabolic diversity of an intermediate salinity hypersaline habitat assessed by metagenomics. *FEMS Microbiol. Ecol.* 88, 623–635. doi: 10.1111/1574-6941.12329
- Fernández, A. B., Vera-Gargallo, B., Sánchez-Porro, C., Ghai, R., Papke, R. T., Rodríguez-Valera, F., et al. (2014b). Comparison of prokaryotic community structure from Mediterranean and Atlantic saltern concentrator ponds by a metagenomic approach. *Front. Microbiol.* 5:196. doi: 10.3389/fmicb.2014.00196
- Foti, M. J., Sorokin, D. Y., Zacharova, E. E., Pimenov, N. V., Kuenen, J. G., and Muyzer, G. (2008). Bacterial diversity and activity along a salinity gradient in soda lakes of the Kulunda Steppe (Altai, Russia). *Extremophiles* 12, 133–145. doi: 10.1007/s00792-007-0117-7
- Gareeb, A. P., and Setati, M. E. (2009). Assessment of alkaliphilic haloarchaeal diversity in Sua pan evaporator ponds in Botswana. *Afr. J. Biotechnol.* 8, 259–267. doi: 10.5897/AJB2009.000-9046
- Ghai, R., Pašić, L., Fernández, A. B., Martín-Cuadrado, A., Mizuno, C. M., McMahon, K. D., et al. (2011). New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.* 1:135. doi: 10.1038/srep00135
- Gnida, M., Ferner, R., Gremer, L., Meyer, O., and Meyer-Klaucke, W. (2003). A novel binuclear [CuSMo] cluster at the active site of carbon monoxide dehydrogenase: characterization by X-ray absorption spectroscopy. *Biochemistry* 42, 222–230. doi: 10.1021/bi026514n
- Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0
- Grant, S., Grant, W. D., Jones, B. E., Kato, C., and Li, L. (1999). Novel archaeal phylotypes from an East African alkaline saltern. *Extremophiles* 3, 139–145. doi: 10.1007/s007920050109
- Grant, W. D. (2004). Life at low water activity. *Philos. Trans. R. Soc. B Biol. Sci.* 359, 1249–1267. doi: 10.1098/rstb.2004.1502
- Gupta, R. S., Naushad, S., and Baker, S. (2014). Phylogenomic analyses and molecular signatures for the class *Halobacteria* and its two major clades: a proposal for division of the class *Halobacteria* into an emended order *Halobacteriales* and two new orders, *Haloferacales* ord. nov. and *Natrialbales* ord. nov. *Int. J. Syst. Evol. Microbiol.* 65, 1050–1069. doi: 10.1099/ijs.0.070136-0
- Haft, D. H., Loftus, B. J., Richardson, D. L., Yang, F., Eisen, J. A., Paulsen, I. T., et al. (2001). TIGRFAMS: a protein family resource for the functional identification of proteins. *Nucleic Acids Res.* 29, 41–43. doi: 10.1093/nar/29.1.41
- Hagedoorn, P. L., Chen, T., Schröder, I., Piersma, S. R., de Vries, S., and Hagen, W. R. (2005). Purification and characterization of the tungsten enzyme aldehyde: ferredoxin oxidoreductase from the hyperthermophilic denitrifier *Pyrobaculum aerophilum*. *J. Biol. Inorg. Chem.* 10, 259–269. doi: 10.1007/s00775-005-0637-5
- Huang, Y., Gilna, P., and Li, W. (2009). Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* 25, 1338–1340. doi: 10.1093/bioinformatics/btp161
- Hügler, M., Huber, H., Molyneux, S. J., Vetriani, C., and Sievert, S. M. (2007). Autotrophic CO<sub>2</sub> fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum *Aquificae*: evidence for two ways of citrate cleavage. *Environ. Microbiol.* 9, 81–92. doi: 10.1111/j.1462-2920.2006.01118.x
- Humayoun, S. B., Bano, N., and Hollibaugh, J. T. (2003). Depth distribution of microbial diversity in mono lake, a meromictic soda lake in California. *Appl. Environ. Microbiol.* 69, 1030–1042. doi: 10.1128/AEM.69.2.1030-1042.2003
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Imhoff, J. F. (1995). “Taxonomy and physiology of phototrophic purple bacteria and green sulfur bacteria,” in *Anoxygenic Photosynthetic Bacteria*, eds R. E. Blankenship, M. T. Madigan and C. E. Bauer (Dordrecht: Springer), 1–15.
- Inoue, K., Ono, H., Abe-Yoshizumi, R., Yoshizawa, S., Ito, H., Kogure, K., et al. (2013). A light-driven sodium ion pump in marine bacteria. *Nat. Commun.* 4:1678. doi: 10.1038/ncomms2689
- Isachenko, B. L. (1951). “Chloride, sulfate and soda lakes of the Kulunda steppe and the biogenic process in them,” in *Selected Works, Vol. 2* (Leningrad: Academy of Sciences USSR), 143–162.
- Johnsen, U., Selig, M., Xavier, K. B., Santos, H., and Schönheit, P. (2001). Different glycolytic pathways for glucose and fructose in the halophilic archaeon *Halococcus saccharolyticus*. *Arch. Microbiol.* 175, 52–61. doi: 10.1007/s002030000237
- Jones, B. F., Eugster, H. P., and Rettig, S. L. (1977). Hydrochemistry of the Lake Magadi basin, Kenya. *Geochim. Cosmochim. Acta* 41, 53–72. doi: 10.1016/0016-7037(77)90186-7
- Jones, C. M., Graf, D. R., Bru, D., Philippot, L., and Hallin, S. (2013). The unaccounted yet abundant nitrous oxide-reducing microbial community: a potential nitrous oxide sink. *ISME J.* 7, 417–426. doi: 10.1038/ismej.2012.125
- Jormakka, M., Yokoyama, K., Yano, T., Tamakoshi, M., Akimoto, S., Shimamura, T., et al. (2008). Molecular mechanism of energy conservation in polysulfide respiration. *Nat. Struct. Mol. Biol.* 15, 730–737. doi: 10.1038/nsmb.1434
- Keshri, J., Mody, K., and Jha, B. (2013). Bacterial community structure in a semi-arid haloalkaline soil using culture independent method. *Geomicrobiol. J.* 30, 517–529. doi: 10.1080/01490451.2012.737092
- King, C. E. (2013). *Diversity and Activity of Aerobic Thermophilic Carbon Monoxide-Oxidizing Bacteria on Kilauea Volcano, Hawaii*. Ph.D. thesis, Louisiana State University.
- King, G. M. (2015). Carbon monoxide as a metabolic energy source for extremely halophilic microbes: implications for microbial activity in Mars regolith. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4465–4470. doi: 10.1073/pnas.1424989112
- King, G. M., and Weber, C. F. (2007). Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* 5, 107–118. doi: 10.1038/nrmicro1595
- Kompantseva, E. I., Bryantseva, I. A., Komova, A. V., and Namsaraev, B. B. (2007). The structure of phototrophic communities of soda lakes of the southeastern Transbaikalian Region. *Microbiology* 76, 211–219. doi: 10.1134/S0026261707020130
- Kompantseva, E. I., Komova, A. V., and Sorokin, D. Y. (2010). Communities of anoxygenic phototrophic bacteria in the saline soda lakes of the Kulunda Steppe (Altai region). *Microbiology* 79, 89–95. doi: 10.1134/S0026261710010121

- Kwon, S. K., Kim, B. K., Song, J. Y., Kwak, M. J., Lee, C. H., Yoon, J. H., et al. (2013). Genomic makeup of the marine flavobacterium *Nonlabens (Donghaeana) dokdonensis* and identification of a novel class of rhodopsins. *Genome Biol. Evol.* 5, 187–199. doi: 10.1093/gbe/evs134
- Lanzen, A., Simachew, A., Gessesse, A., Chmolsowska, D., Jonassen, I., and Øvreås, L. (2013). Surprising prokaryotic and eukaryotic diversity, community structure and biogeography of Ethiopian soda lakes. *PLoS ONE* 8:e72577. doi: 10.1371/journal.pone.0072577
- Lassmann, T., and Sonnhammer, E. L. (2005). Kalign—an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* 6:298. doi: 10.1186/1471-2105-6-298
- Lauro, F. M., DeMaere, M. Z., Yau, S., Brown, M. V., Ng, C., Wilkins, D., et al. (2011). An integrative study of a meromictic lake ecosystem in Antarctica. *ISME J.* 5, 879–895. doi: 10.1038/ismej.2010.185
- León, M. J., Fernández, A. B., Ghai, R., Sánchez-Porro, C., Rodríguez-Valera, F., and Ventosa, A., (2014). From metagenomics to pure culture: isolation and characterization of the moderately halophilic bacterium *Spiribacter salinus* gen. nov., sp. nov. *Appl. Environ. Microbiol.* 80, 3850–3857. doi: 10.1128/AEM.00430-14
- Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033
- Liu, X., Gao, C., Zhang, A., Jin, P., Wang, L., and Feng, L. (2008). The *nos* gene cluster from gram-positive bacterium *Geobacillus thermodenitrificans* NG80-2 and functional characterization of the recombinant NosZ. *FEMS Microbiol. Lett.* 289, 46–52. doi: 10.1111/j.1574-6968.2008.01362.x
- Llorens-Marès, T., Yooseph, S., Goll, J., Hoffman, J., Vila-Costa, M., Borrego, C. M., et al. (2015). Connecting biodiversity and potential functional role in modern euxinic environments by microbial metagenomics. *ISME J.* 9, 1648–1661. doi: 10.1038/ismej.2014.254
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42(Database issue):D490–D495. doi: 10.1093/nar/gkt1178
- López-Pérez, M., Ghai, R., Leon, M. J., Rodríguez-Olmos, Á., Copa-Patiño, J. L., Soliveri, J., et al. (2013). Genomes of “*Spiribacter*,” a streamlined, successful halophilic bacterium. *BMC Genomics* 14:787. doi: 10.1186/1471-2164-14-787
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.0955
- Lynch, M. D., and Neufeld, J. D. (2015). Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* 13, 217–229. doi: 10.1038/nrmicro3400
- Marmur, J. (1961). A procedure for isolation of DNA from microorganisms. *J. Mol. Biol.* 3, 208–214. doi: 10.1016/S0022-2836(61)80047-8
- Martin-Cuadrado, A. B., Ghai, R., Gonzaga, A., and Rodríguez-Valera, F. (2009). CO dehydrogenase genes found in metagenomic fosmid clones from the deep Mediterranean Sea. *Appl. Environ. Microbiol.* 75, 7436–7444. doi: 10.1128/AEM.01283-09
- Martin-Cuadrado, A. B., Pašaić, L., and Rodríguez-Valera, F. (2015). Diversity of the cell-wall associated genomic island of the archaeon *Haloquadratum walsbyi*. *BMC Genomics* 16:603. doi: 10.1186/s12864-015-1794-8
- Martínez-García, M., Santos, F., Moreno-Paz, M., Parro, V., and Antón, J. (2014). Unveiling viral–host interactions within the ‘microbial dark matter’. *Nat. Commun.* 5:4542. doi: 10.1038/ncomms5542
- Melack, J. M. (1981). Photosynthetic activity of phytoplankton in tropical African soda lakes. *Hydrobiologia* 81, 71–85. doi: 10.1007/BF00048707
- Mesbah, N. M., Abou-El-El, S. H., and Wiegel, J. (2007). Novel and unexpected prokaryotic diversity in water and sediments of the alkaline, hypersaline lakes of the Wadi An Natrun, Egypt. *Microbiol. Ecol.* 54, 598–617. doi: 10.1007/s00248-006-9193-y
- Milford, A. D., Achenbach, L. A., Jung, D. O., and Madigan, M. T. (2000). *Rhodobaca bogoriensis* gen. nov. and sp. nov., an alkaliphilic purple nonsulfur bacterium from African Rift Valley soda lakes. *Arch. Microbiol.* 174, 18–27. doi: 10.1007/s002030000166
- Mongodin, E. F., Nelson, K. E., Daugherty, S., Deboy, R. T., Wister, J., Khouri, H., et al. (2005). The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic Bacteria and Archaea. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18147–18152. doi: 10.1073/pnas.0509073102
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, 182–185. doi: 10.1093/nar/gkm321
- Mwirichia, R., Cousin, S., Muigai, A. W., Boga, H. I., and Stackebrandt, E. (2011). Bacterial diversity in the haloalkaline Lake Elmteita, Kenya. *Curr. Microbiol.* 62, 209–221. doi: 10.1007/s00284-010-9692-4
- Narasingarao, P., Podell, S., Ugalde, J. A., Brochier-Armanet, C., Emerson, J. B., Brocks, J. J., et al. (2012). *De novo* metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J.* 6, 81–93. doi: 10.1038/ismej.2011.78
- Ochsenreiter, T., Pfeifer, F., and Schleper, C. (2002). Diversity of Archaea in hypersaline environments characterized by molecular-phylogenetic and cultivation studies. *Extremophiles* 6, 267–274. doi: 10.1007/s00792-001-0253-4
- Oren, A. (1994). The ecology of the extremely halophilic Archaea. *FEMS Microbiol. Rev.* 13, 415–439. doi: 10.1111/j.1574-6976.1994.tb00060.x
- Oren, A. (2013). Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front. Microbiol.* 4:315. doi: 10.3389/fmicb.2013.00315
- Oren, A. (2014). Taxonomy of halophilic Archaea: current status and future challenges. *Extremophiles* 18, 825–834. doi: 10.1007/s00792-014-0654-9
- Pagalang, E., Wang, H., Venables, M., Wallace, A., Grant, W. D., Cowan, D. A., et al. (2009). Microbial biogeography of six salt lakes in inner Mongolia, China, and a salt lake in Argentina. *Appl. Environ. Microbiol.* 75, 5750–5760. doi: 10.1128/AEM.00040-09
- Park, B. H., Karpnits, T. V., Syed, M. H., Leuze, M. R., and Uberbacher, E. C. (2010). CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* 20, 1574–84. doi: 10.1093/glycob/cwq106
- Park, S., Akira, Y., Kazuhiro, K. (2014). “The family *Rhodothermaceae*,” in *The Prokaryotes—Other Major Lineages of Bacteria and the Archaea*, eds E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin Heidelberg: Springer-Verlag), 849–856.
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2014). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Petitjean, C., Deschamps, P., López-García, P., and Moreira, D. (2015). Rooting the domain Archaea by phylogenomic analysis supports the foundation of the new kingdom *Protearchaeota*. *Genome Biol. Evol.* 7, 191–204. doi: 10.1093/gbe/evu274
- Rees, H. C., Grant, W. D., Jones, B. E., and Heaphy, S. (2004). Diversity of Kenyan soda lake alkaliphiles assessed by molecular methods. *Extremophiles* 8, 63–71. doi: 10.1007/s00792-003-0361-4
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2
- Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437. doi: 10.1038/nature12352
- Rodríguez-Valera, F., Ventosa, A., Juez, G., and Imhoff, J. F. (1985). Variation of environmental features and microbial populations with salt concentrations in a multi-pond saltern. *Microbiol. Ecol.* 11, 107–115. doi: 10.1007/BF02010483
- Say, R. F., and Fuchs, G. (2010). Fructose 1, 6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* 464, 1077–1081. doi: 10.1038/nature08884
- Sharon, I., and Banfield, J. F. (2013). Genomes from metagenomics. *Science* 342, 1057–1058. doi: 10.1126/science.1247023
- Siebers, B., Brinkmann, H., Dörr, C., Tjaden, B., Lilie, H., van der Oost, J., et al. (2001). Archaeal fructose-1, 6-bisphosphate aldolases constitute a new family of archaeal type class I aldolase. *J. Biol. Chem.* 276, 28710–28718. doi: 10.1074/jbc.M103447200
- Siebers, B., and Schönheit, P. (2005). Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. *Curr. Opin. Microbiol.* 8, 695–705. doi: 10.1016/j.mib.2005.10.014
- Simachew, A., Lanzén, A., Gessesse, A., and Øvreås, L., (2015). Prokaryotic community diversity along an increasing salt gradient in a soda ash concentration pond. *Microb. Ecol.* 71, 326–338. doi: 10.1007/s00248-015-0675-7

- Sorokin, D. Y., Abbas, B., Geleijnse, M., Pimenov, G. N., Sukhacheva, M. V., and van Loosdrecht, M. C. (2015b). Methanogenesis at extremely haloalkaline conditions in soda lakes of Kulunda Steppe (Altai, Russia). *FEMS Microbiol. Ecol.* 91:fiv016. doi: 10.1093/femsec/fiv016
- Sorokin, D. Y., Banciu, H. L., and Muyzer, G. (2015a). Functional microbiology of soda lakes. *Curr. Opin. Microbiol.* 25, 88–96. doi: 10.1016/j.mib.2015.05.004
- Sorokin, D. Y., Berben, T., Melton, E. D., Overmars, L., Vavourakis, C. D., and Muyzer, G. (2014). Microbial diversity and biogeochemical cycling in soda lakes. *Extremophiles* 18, 791–809. doi: 10.1007/s00792-014-0670-9
- Sorokin, D. Y., Janssen, A. J., and Muyzer, G. (2012). Biodegradation potential of halo (alkali) philic prokaryotes. *Crit. Rev. Environ. Sci. Technol.* 42, 811–856. doi: 10.1080/10643389.2010.534037
- Sorokin, D. Y., Kublanov, I. V., Gavrilov, S. N., Rojo, D., Roman, P., Golyshin, P. N., et al. (2015c). Elemental sulfur and acetate can support life of a novel strictly anaerobic haloarchaeon. *ISME J.* 9, 240–252. doi: 10.1038/ismej.2015.79
- Sorokin, D. Y., Toshchakov, S. V., Kolganova, T. V., and Kublanov, I. V. (2015d). Halo(natrono)archaea isolated from hypersaline lakes utilize cellulose and chitin as growth substrates. *Front. Microbiol.* 6:942. doi: 10.3389/fmicb.2015.00942
- Sorokin, D. Y., Tourova, T. P., Galinski, E. A., Belloch, C., and Tindall, B. J. (2006). Extremely halophilic denitrifying bacteria from hypersaline inland lakes *Halovibrio denitrificans* sp. nov. and *Halospina denitrificans* gen. nov., sp. nov., and evidence that the genus name *Halovibrio* (Fendrich 1989) with the type species *H. variabilis* should be associated with DSM 3050. *Int. J. Syst. Evol. Microbiol.* 56, 379–388. doi: 10.1099/ijs.0.63964-0
- Sorokin, D. Y., Tourova, T. P., Kuznetsov, B. B., Bryantseva, I. A., and Gorlenko, V. M. (2000). *Roseinatronobacter thiooxidans* gen. nov., sp. nov., a new alkaliphilic aerobic bacteriochlorophyll a—containing bacterium isolated from a soda lake. *Microbiology* 69, 75–82. doi: 10.1007/BF02757261
- Tamura, K., Strecher, G., Peterson, D., Filipowski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. doi: 10.1186/1471-2105-4-41
- Tindall, B. J. (1988). “Prokaryotic life in the alkaline, saline, athalassic environment,” in *Halophilic Bacteria*, ed F. Rodriguez-Valera (Boca Raton, FL: CRC Press), 31–67.
- Tourova, T. P., Grechnikova, M. A., Kuznetsov, B. B., and Sorokin, D. Y. (2014). Phylogenetic diversity of bacteria in soda lake stratified sediments. *Microbiology* 83, 869–879. doi: 10.1134/S0026261714060186
- Tourova, T. P., Kovaleva, O. L., Bumazhkin, B. K., Patutina, E. O., Kuznetsov, B. B., Bryantseva, I. A., et al. (2011). Application of ribulose-1, 5-bisphosphate carboxylase/oxygenase genes as molecular markers for assessment of the diversity of autotrophic microbial communities inhabiting the upper sediment horizons of the saline and soda lakes of the Kulunda Steppe. *Microbiology* 80, 812–825. doi: 10.1134/S0026261711060221
- Tremblay, J., Singh, K., Fern, A., Kirton, E. S., He, S., Woyke, T., et al. (2015). Primer and platform effects on 16S rRNA tag sequencing. *Front. Microbiol.* 6:771. doi: 10.3389/fmicb.2015.00771
- Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M. et al. (2004). Community structure and metabolism through reconstruction of microbial genomes. *Nature* 428, 37–43. doi: 10.1038/nature02340
- Ugalde, J. A., Podell, S., Narasingarao, P., Allen, E. E. (2011). Xenorhodopsins, an enigmatic new class of microbial rhodopsins horizontally transferred between archaea and bacteria. *Biol. Direct* 6:52. doi: 10.1186/1745-6150-6-52
- Urios, L., Agogue, H., Lesongeur, F., Stackebrandt, E., and Lebaron, P. (2006). *Balneola vulgaris* gen. nov., sp. nov., a member of the phylum *Bacteroidetes* from the north-western Mediterranean Sea. *Int. J. Syst. Evol. Microbiol.* 56, 1883–1887. doi: 10.1099/ijs.0.64285-0
- Vaisman, N., and Oren, A. (2009). *Salisaeta longa* gen. nov., sp. nov., a red, halophilic member of the *Bacteroidetes*. *Int. J. Syst. Evol. Microbiol.* 59, 2571–2574. doi: 10.1099/ijs.0.010892-0
- Wang, Y. X., Li, Y. P., Liu, J. H., Xiao, W., Lai, Y. H., Li, Z. Y., et al. (2013). *Gracilimonas mengyeensis* sp. nov., a moderately halophilic bacterium isolated from a salt mine in Yunnan, south-western China. *Int. J. Syst. Evol. Microbiol.* 63, 3989–3993. doi: 10.1099/ijs.0.052043-0
- Wood, D., Setubal, J., Kaul, R., Monks, D., Kitajima, J., Okura, V., et al. (2001). The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* 294, 2317–2323. doi: 10.1126/science.1066804
- Yoon, J. H., Kang, S. J., Jung, Y. T., and Oh, T. K. (2009). *Psychroflexus salinarum* sp. nov., isolated from a marine solar saltern. *Int. J. Syst. Evol. Microbiol.* 59, 2404–2407. doi: 10.1099/ijs.0.008359-0
- Youssef, N. H., Savage-Ashlock, K. N., McCully, A. L., Luedtke, B., Shaw, E. I., Hoff, W. D., et al. (2014). Trehalose/2-sulfotrehalose biosynthesis and glycine-betaine uptake are widely spread mechanisms for osmoadaptation in the *Halobacteriales*. *ISME J.* 8, 636–649. doi: 10.1038/ismej.2013.165

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Vavourakis, Ghai, Rodriguez-Valera, Sorokin, Tringe, Hugenholtz and Muyzer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.