Accepted Manuscript

Title: Genomic comparison between members of the *Salinibacteraceae* family, and description of a new species of *Salinibacter* (*Salinibacter altiplanensis* sp. nov.) isolated from high altitude hypersaline environments of the Argentinian Altiplano

Authors: Tomeu Viver, Luis Orellana, Pedro González-Torres, Sara Díaz, Mercedes Urdiain, María Eugenia Farías, Vladimir Benes, Peter Kaempfer, Azadeh Shahinpei, Mohammad Ali Amoozegar, Rudolf Amann, Josefa Antón, Konstantinos T. Konstantinidis, Ramon Rosselló-Móra

PII: S0723-2020(18)30025-0
DOI: https://doi.org/10.1016/j.syapm.2017.12.004
Reference: SYAPM 25892

To appear in:

Received date: 22-11-2017
Revised date: 22-12-2017
Accepted date: 23-12-2017

Please cite this article as: Tomeu Viver, Luis Orellana, Pedro González-Torres, Sara Díaz, Mercedes Urdiain, María Eugenia Farías, Vladimir Benes, Peter Kaempfer, Azadeh Shahinpei, Mohammad Ali Amoozegar, Rudolf Amann, Josefa Antón, Konstantinos T.Konstantinidis, Ramon Rosselló-Móra, Genomic comparison between members of the Salinibacteraceae family, and description of a new species of Salinibacter (Salinibacter altiplanensis sp.nov.) isolated from high altitude hypersaline environments of the Argentinian Altiplano, Systematic and Applied Microbiology https://doi.org/10.1016/j.syapm.2017.12.004

Genomic comparison between members of the *Salinibacteraceae* family, and description of a new species of *Salinibacter* (*Salinibacter altiplanensis* sp. nov.) isolated from high altitude hypersaline environments of the Argentinian Altiplano.

Tomeu Viver[1], Luis Orellana[2], Pedro González-Torres[3], Sara Díaz[1], Mercedes Urdiain[1], María Eugenia Farías[4], Vladimir Benes[5], Peter Kaempfer[6], Azadeh Shahinpei[7], Mohammad Ali Amoozegar[7], Rudolf Amann[8], Josefa Antón[3], Konstantinos T. Konstantinidis[2], Ramon Rosselló-Móra[1]

Affiliations:

[1] Marine Microbiology Group, Department of Ecology and Marine Resources, Mediterranean Institute for Advanced Studies (IMEDEA, CSIC-UIB), Esporles, Spain

[2] School of Civil & Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

[3] Department of Physiology, Genetics and Microbiology, University of Alicante, Alicante, Spain

[4] Laboratorio de Investigaciones Microbiológicas de Lagunas Andinas (LIMLA), Planta Piloto de Procesos Industriales Microbiológicos (PROIMI), CCT, CONICET, Tucumán, Argentina

[5] European Molecular Biology Laboratory (EMBL), Genomics Core Facility, Meyerhofstr. 1, D-69117 Heidelberg, Germany

[6] Institut für Angewandte Mikrobiologie, Justus-Liebig-Universität Giessen, IFZ-Heinrich-Buff-Ring, Giessen, Germany

[7] Extremophiles Laboratory, Department of Microbiology, Faculty of Biology and Center of Excellence in Phylogeny of Living Organisms, Collage of Science, University of Tehran, Tehran, Iran

[8] Department of Molecular Ecology, Max-Planck-Institut für Marine Mikrobiologie, Bremen, D-28359, Germany.

Corresponding Author:

Tomeu Viver,
Marine Microbiology Group,
Department of Ecology and Marine Resources,
Mediterranean Institute for Advanced Studies (IMEDEA, CSIC-UIB),
E-07190, Esporles,
Spain
Tel: +34 971611827
Email: tviver@imedea.uib-csic.es

Abstract

The application of tandem MALDI-TOF MS screening with 16S rRNA gene sequencing of selected isolates has been demonstrated to be an excellent approach for retrieving novelty from large-scale culturing. The application of such methodologies in different hypersaline samples allowed the isolation of the culture-recalcitrant *Salinibacter ruber* second phylotype (EHB-2) for the first time, as well as a new species recently isolated from the Argentinian Altiplano hypersaline lakes. In this study, the genome sequences of the different species of the phylum *Rhodothermaeota* were compared and the genetic repertoire along the evolutionary gradient was analyzed together with each intraspecific variability. Altogether, the results indicated an open pan-genome for the family *Salinibacteraceae*, as well as the codification of relevant traits such as diverse rhodopsin genes, CRISPR-Cas systems and spacers, and one T6SS secretion system that could give ecological advantages to an EHB-2 isolate. For the new *Salinibacter* species, we propose the name *Salinibacter altiplanensis* sp. nov. (the designated type strain is AN15^T =CECT 9105^T = IBRC-M 11031^T)

Key Words: *Salinibacter ruber*, halophiles, MALDI-TOF MS, salterns, salt lake, genomic.

Introduction

Salinibacter ruber has been discovered to be the first extreme halophilic member of the domain Bacteria with cell abundances equivalent to most haloarchaea [4]. *Sal. ruber* (for the rationale of the abbreviations used see Materials and Methods) was first observed using culture-independent approaches, which led to the description of *Candidatus* “Salinibacter” that was subsequently formally classified as a new species after isolation in pure culture [3]. *Sal. ruber* was the first species described from a new lineage of extreme halophilic microorganisms, monophyletic with another extremophile (*Rhodothermus marinus*), and which were both loosely affiliated to the *Chlorobium* and *Bacteroidetes* phyla [65]. Extensive research in hypersaline environments led to the isolation and classification of two additional members of the genus *Salinibacter* (*Sal. iranicus* and *Sal. luteus*) from the Iranian salt lake Aran-Bidgol [36], and one new genus *Salisaeta* (*Sat. longa* [68]) that originated from a mixture of water from the Dead Sea and the Red Sea. Recently, the lineage comprising the three genera has been classified as a single phylum *Rhodothermaeota*, and the Iranian species were reclassified as the new genus *Salinivenus* [42] due to their phylogenetic distance from the type species of *Salinibacter*.

Since its discovery, numerous members of the species *Sal. ruber* have been isolated from diverse and distant environments (e.g. [2, 63]). This species was shown to be formed by two closely related phylotypes (EHB-1 and EHB-2) coexisting in the same environment [4]. However, in all cases, the isolates corresponded to the most abundant phylotype EHB-1, and during almost 20 years of research the members of the second phylotype EHB-2 escaped isolation. Representatives of this species have been isolated from several hypersaline spots

around the world even from distant locations and different altitudes, such as Mediterranean coastal salterns and Peruvian Altiplano salterns [63]. Furthermore, sequences related to the phylotypes have been detected in several culture-independent surveys from different locations, such as Tuz Lake in Turkey [43], lakes in the Tibetan plateau [74], the Argentinian Pampa [18], or Lake Tyrrell in Australia [54].

High-throughput culturing in tandem with MALDI-TOF MS screening and 16S rRNA identification of isolates from hypersaline systems is a robust strategy for retrieving rare taxa from environmental samples [70]. Using this strategy, the pure cultures of two strains corresponding to the *Sal. ruber* EHB-2 phylotype are described in this current study together with members of a new species of *Salinibacter* thriving in hypersaline lakes of the Argentinian Altiplano. The genomes of these strains, together with the type strains of *Salinivenus*, were sequenced and compared with the available *Rhodothermaeota* genomes to reveal new genomic features of this extremely halophilic lineage.

Materials and Methods

Strains and name abbreviations

Reference strains of *Sal. ruber* M31^T and M8 were obtained from our strain collection, and the type strains of *Slv. iranica* (CB7^T) and *Slv. lutea* (DGO^T) were provided by the co-author M. Amoozegar from his collection. In order to simplify the identification of the names, a three letter abbreviation was used, which was already commonly used for *Halobacteria* in accordance with the recommendation made by the International Committee on Systematic Bacteriology, Subcommittee on Taxonomy of *Halobacteriaceae* [46] as: *Salinibacter* = *Sal*.; *Salinivenus* = *Slv*.; *Salisaeta* = *Sat*.; and *Rhodothermus* = *Rho*.

Sample processing and strains studied

Two different athalassohaline salt lakes (Ojo Rojo in Antofalla and Salar de Llullialliaco), both located in the Argentinian Altiplano at altitudes above 3,600 m, were sampled in February 2011 (Table 1). The salinity of the brines was 34% and the pH was 7. Cultures were obtained using salt water medium (SW) at a salt concentration of 25% [61]. The isolated strains were screened by MALDI-TOF MS (Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry) using whole cell biomass, as previously published [70]. An in-house MALDI-TOF profile database was used to generate a dendrogram and detect new *Salinibacter* isolates. 16S rRNA gene PCR amplification of the isolates was performed as previously published [70]. The two strains ST67 and SP273 had been isolated previously in a survey from the Trinitat (Tarragona, Spain) and Santa Pola (Alicante, Spain) solar salterns, respectively [70]. The type strains of *Slv. iranica* and *Slv. lutea* (CB7^T and DGO^T, respectively) were added to the genome sequencing strategy. For the global comparisons, the genomes of the already characterized *Sal. ruber* M31^T and M8 strains, and *Salisaeta longa* S4-4^T were incorporated (Table 1).

DNA extraction and genome assembly analysis

DNA extraction was performed as detailed by Urdiain et al. [67] and different methods of sequencing were used: Illumina Miseq (PE 300x2), Illumina Hiseq (PE 100x2), Roche GS FLX and Pacific Biosciences PacBio RS (Table 2). Illumina and Roche GS FLX reads were trimmed with a PHRED score quality threshold of 20 using SolexaQA v3.1.4 [14]. Different assembly softwares were used for each sequencing platform. Genomes sequenced by Illumina MiSeq were assembled using IDBA v1.1.1 [48] and by Illumina HiSeq using Velvet v1.0.13 [73]. A hybrid assembly methodology was used for assembling strain ST67, sequenced by Roche and PacBio: trimmed sequences from Roche were assembled using SPADES v.3.1.1 [6] and then ordered by using the long read information from the PacBio backbone using SSPACE-LongReads v1.1 [9]. Gene prediction from assembled contigs was conducted by using GeneMark.hmm with default parameters [8], and functional annotation was based on protein level searches against NCBI databases with Blast2Go v3.0.10 [13]. The annotations were compared with the RAST annotation, and metabolic pathways were analyzed using KAAS-KEGG [41]. CRISPR spacers were predicted in the genomes using CRISPRfinder [25].

Tree reconstructions based on rRNA and housekeeping genes

16S rRNA gene sequences were retrieved from the genomes and the alignments, and tree reconstructions were performed using the ARB software package version 5.5 [32]. The new sequences were added to the reference dataset LTP115 [72] and aligned using the SINA v1.2.12 tool (SILVA Incremental Aligner [55]) implemented within the ARB software package. Final alignments were manually improved following the reference alignment in ARB-editor. Sequences were used to reconstruct *de novo* trees using the neighbor-joining algorithm. 23S rRNA gene sequences and multilocus sequence analysis (MLSA) with 29 single-copy genes were also extracted from assembled genomes. The 23S rRNA genes were added to the LSURef 115 SILVA dataset, and selected sequences were aligned using the SINA aligner implemented in ARB-editor. The MLSA genes selected in this study were the same as those used in the revised phylogeny of *Bacteroidetes* [42]. MLSA genes were aligned individually using MUSCLE v3.8.31 [21] and were concatenated posteriorly. The neighbor-joining (NJ) [64] and RaxML v8.2.0 [66] algorithms were used for phylogenetic reconstructions as implemented in ARB. Tree reconstructions with NJ were performed using the Jukes Cantor correction, and RaxML reconstructions with the GTRGAMMA correction.

Core and pan-genome analysis; phylogenetic reconstruction

Predicted protein sequences were compared using an all-versus-all BLAST v2.2.28 [1] with available reference sequences in order to identify the shared reciprocal best matches in all pairwise genome comparisons using a 50% sequence similarity cut-off and over 50% or more of the query sequence length. All proteins shared between all sequenced genomes were aligned using MUSCLE v3.8.31 [21]. The concatenated and aligned orthologous genes were used to build phylogenetic trees in RAxML v8.2.0 [66]. The variable genes were defined as those

present in two or more genomes but not in all genomes. The presence or absence of variable genes was used to cluster the genomes with the Euclidian distance using the Ggplot2 package from R [71].

Synteny regions

Assembled contigs were sorted using Mauve Contig Mover [58]. Maximum unique matches with a minimum cluster length of 20 nucleotides were calculated using the NUCmer function from the MUMer package v3.0. Regions of synteny were identified by the visualization of the data using MUMmerplot [31].

ANI and AAI computation

The average nucleotide identity (ANI) and the average amino acid identity (AAI) between all genomes were determined according to Konstantinidis and Tiedje [28] using the webserver available at http://enve-omics.gatech.edu/ [60]. AAI values were calculated using core-genome genes after comparing all the genomes from the family *Salinibacteraceae* and adding the *Rhodothermus marinus* DSM 4252^T genome.

Physiological and biochemical tests

The following tests were performed as outlined by Cowan and Steel's Manual for the Identification of Medical Bacteria [7]: catalase, oxidase, anaerobic growth in the presence of arginine and DMSO; hydrolysis of Tween 20 and DNA, casein, gelatin and starch; arginine dihydrolase, lysine decarboxylase, tryptophanase and ornithine decarboxylase activities; methyl-red and Voges-Proskauer reactions; H_2S production, gas formation with nitrate. Physiological tests were performed with SW or MGM broth or agar medium. Broth cultures were incubated in an orbital incubator at 250 r.p.m. Growth kinetics were followed using OD_{600} *versus* time graphs produced from measurements in an Eppendorf biophotometer. Optimum temperatures were tested between 20 to 60 ºC. The pH optimum was determined between 5 and 9. The pH of the medium was adjusted by 50 mM MES (pH 5-6.5), HEPES (pH 7-8) and CHES (8.5-9) buffers. Pigments were obtained by using methanol/acetone (1:1, v/v) as an extraction solution after the addition of cell pellets. Acid production from carbohydrates (0.1%, w/v) was determined in unbuffered MGM broth (pH 7.5) by measuring the initial and final pH of the medium. A culture was considered positive for acid production if the pH decreased by at least 1 pH unit. The test was also repeated with the same medium and 0.001% phenol red pH indicator. To test for carbon source utilization (1%, w/v) peptone was omitted from the MGM broth (pH 7.5) and the yeast extract concentration was reduced to 0.1 g L^{-1}. The absorption spectra of the strains were obtained using a HITACHI U-2900 spectrophotometer, as previously described [3]. The whole cell fatty acid composition was determined by following the standard protocol of the Microbial Identification System (MIDI; Sherlock version 6.1). The extracts obtained were analyzed by using a GC (HP6890A; Hewlett Packard).

Microscopy

Cell morphology and motility of the organisms were observed under a Zeiss Axio Imager A1 optical microscope. For photography, drops of exponentially growing SW broth cultures were used directly without fixing. Gram stains were prepared following the method of Dussault [20]. Catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) and probe stringency optimization were performed following protocols previously reported [53].

Results and Discussion

Isolation of new representatives of the genus Salinibacter

The *Sal. ruber* species, initially discovered and classified as *Candidatus*, can be regarded as a rare case where the isolation of a pure culture of a representative of a candidate taxon was achieved very shortly after its discovery [45]. However, from the two phylotypes detected by FISH [4], only the most abundant EHB-1 had been cultured, and it is already represented by a large collection of isolates (e.g. [2, 63]), whereas the second EHB-2 phylotype has escaped cultivation. However, the MALDI-TOF / 16S rRNA gene sequencing approach [70] for 4,200 isolates where *Sal. ruber* EHB-1 dominated, finally yielded two strains, ST67 and SP273 (Table 1), from the Trinitat (Delta del Ebro) and Santa Pola (Alicante) salterns, respectively, which were representative of EHB-2. Both strains formed a single lineage according to 16S rRNA gene phylogenetic analysis, which was different from that of the *Sal. ruber* hitherto cultured (Figure 1A), and they affiliated closely with the EHB-2 phylotype of *Sal. ruber* [4, 70]. The two strains showed a 16S rRNA gene identity value with type strain M31^T of 97.4% and 98.0%, and 98.2% between themselves (Supplementary Table S1). These two isolates were the only members of the EHB-2 phylotype among the 1,613 strains identified as members of *Sal. ruber* species.

In this current study, the same MALDI-TOF / 16S rRNA tandem gene sequencing approach was applied to two unexplored hypersaline high altitude lakes of the Argentinian Altiplano (Ojo Rojo in Antofalla and Salar de Llullialliaco located at 4,000 and 3,677 m a.s.l., respectively; Table 1). From these samples, 58 strains were isolated that corresponded to 23 Archaea and 35 Bacteria. The bacterial strains formed a single cluster that was closely related but distinct from *Sal. ruber* (Supplementary Figure S1) and, unexpectedly, no representative of *Sal. ruber* could be isolated. Three isolates, AN4 and AN15^T from Antofalla and LL19 from Llullialliaco, were used for further studies. The results indicated that these isolates were members of the same *Sal. ruber* lineage, but they were affiliated as a distinct branch, with 96.5% – 96.8% 16S rRNA gene identity with the type strain M31^T (Figure 1A, Supplementary Table S1). In addition, the strains loosely affiliated with the members of the neighboring genera *Salinivenus* and *Salisaeta* with sequence identity values ranging from 91.6% to 92.1% and 87.9% to 88.0%, respectively. Conspicuously, a culture-independent survey from the same samples showed the presence of different lineages of *Salinibacter*, but none corresponded to the phylotype formed by the three isolates [40], which indicated that the latter phylotype may not have represented the most

abundant members of the *Salinibacteraceae* lineage in the samples studied, although it was the only cultivable member of the family obtained under the standard laboratory conditions used. On the other hand, this phylotype may be characteristic of the South American region, since similar sequences have been retrieved by amplicon sequencing from La Pampa province in central Argentina [18].

Comparisons between the sequenced genomes of the Rhodothermaeota

The two new representatives of the EHB-2 lineage, ST67 and SP273, three representatives of the Argentinian isolates AN4 and AN15^T and LL19, and the two type strains of *Slv. iranica* (CB7^T) and *Slv. lutea* (DGO^T) were sequenced (Table 2), and the genomes obtained were compared with the available genomes of *Sal. ruber sensu stricto* and the *Sat. longa* (S4-4^T) and *Rho. marinus* (DSM 4252^T) type strains. All the genomes, except those of *Slv. iranica* (CB7^T) and SP273, were obtained in fewer than 100 contigs. In all cases, the sequencing coverage was over 36-fold, which provided almost complete genome representation [33]. It was only possible to close the genome of strain ST67, which was obtained in a single chromosome of 3.5 Mb and two plasmids of 0.18 Mb and 0.09 Mb, respectively. The smallest genomes with 3.41 Mb and 3.40 Mb corresponded to *Slv. iranica* (CB7^T) and *Sat. longa* (S4-4^T), respectively. On the other hand, the largest sequenced genome was the EHB-2 representative SP273 with 4.06 Mb. In general, the G+C mol% of the genomes was well within the range calculated for the members of the lineage and ranged between 64.2% for the Argentinian strains and 66.2% for EHB-2 representative ST67. The three Argentinian isolates always exhibited the lowest G+C mol% of the collection. The number of predicted coding sequences (CDS) was similar in all cases and ranged between 2,931 CDS in *Slv. iranica* (CB7^T) to 3,362 CDS in SP273. Between 32.3% (for CB7^T) and 38.7% (for SP273) of the detected CDS corresponded to hypothetical proteins. In all cases, a single rRNA operon was detected and it always followed the same order 16S – 23S – 5S.

The reconstructed phylogenies based on the 23S rRNA gene (Figure 1B), a concatenate of 29 conserved, single-copy orthologous gene products [42] (Figure 1C), and the concatenate of the core genome (Figure 1D) showed nearly identical topologies. The 16S rRNA gene and the complete core genome better resolved the phylogenetic divergence between EHB-1 and EHB-2 phylotypes where SP273 and ST67 affiliated as a distinct lineage from M31^T and M8, indicating a slight common evolutionary divergence [4]. The three strains isolated from the Argentinian Altiplano (i.e. AN15^T, AN4 and LL19) formed a single branch loosely affiliated with the representatives of *Sal. ruber*. The representatives of the genus *Salinivenus* (i.e. *Slv. iranica* (CB7^T) and *Slv. lutea* (DGO^T)) were always distantly affiliated to *Sal. ruber* and the Argentinian strains. *Sat. longa* (S4-4^T) was the most divergent strain within the *Salinibacteraceae* family. The topology of the different trees, and the sequence identity between the clades were in accordance with the distinct species and genus classification (<98.7% for species and <94.5%

for genus [72]), indicating that the Argentinian strains could represent a new species of the genus *Salinibacter*.

Both strains ST67 and SP273, despite a clear phylogenetic divergence, could undoubtedly be identified as members of the same species by their high ANI and AAI values above 98.2% and 99% (sharing more than 66% of genes), respectively (Supplementary Table S2), with the type strain of the species M31^T [24, 57, 62]. Similarly, all Argentinian strains shared 98.3% and 99% (sharing more than 75% of genes) ANI and AAI values, respectively. In addition, both type strains of the two *Salinivenus* species shared 96.3% and 98.49% ANI and AAI values, respectively. On the other hand, the ANI and AAI values between *Sal. ruber* and the Argentinian strains were always below 85.8% and 85.4%. The AAI values between *Sal. ruber* and *Salinivenus* were always below 71%, which was close to the plausible value for discriminating genera using whole genome comparisons [27], and supporting the classification of the genus *Salinivenus* [42]. Both parameters decreased in parallel with the evolutionary distance measured by 16S rRNA gene identities, as well as the percentages of shared genes (Figure 2). Differently to the fast ANI decrease along the evolutionary divergence, AAI decreased linearly with the evolutionary distance and with the percentage of shared genes, in accordance with that already reported as a general trend [27]. In all cases, the ANI values within each single major lineage were in accordance with their consideration as distinct species [24, 57], and the AAI values correlated strongly with the different classified categories [27].

Gene content and other features of the sequenced genomes

The core genome between the three genera *Salinibacter*, *Salinivenus* and *Salisaeta*, and the single genome of *Rho. marinus* (DSM 4252^T) shared 923 genes that represented between 28% and 32% of the respective genomes. This core gene set was mainly formed by genes of the central metabolism and only 33 genes (3.6% of the core) were annotated as hypothetical conserved proteins (Supplementary Spreadsheet). On the other hand, the complete set of auxiliary genes (i.e. not present in all genomes) was approximately 9,000 (Figure 3; and Supplementary Spreadsheet). The tendency of the core genome reduction by adding the different members of the phylum was asymptotic, whereas the pan-genome still seemed to be steadily increasing. Similar trends were observed at the single species level (Supplementary Figure S2) pointing to an open-genome trend [37]. This was not surprising as we had already demonstrated that the *Sal. ruber* genomes and metabolomes were very diverse [2].

Three groups of strains were detected that would represent single species, each with multiple strains. These were *Sal. ruber* with M31^T, M8, ST67 and SP273 (Figures 1 and 4A); the Argentinian isolates AN15^T, AN4 and LL19 (Figures 1 and 4B); and the pair of type strains of the genus *Salinivenus* (i.e. *Slv. iranica* CB7^T and *Slv. lutea* DGO^T; Figures 1 and 4C). Although the last two strains shared an ANI value >96% that could be an indication of being potentially a single species, their distinct phenotypes [36] justified their maintenance as distinct species. In

these three cases, the core genome of each species ranged between 2,416 and 2,584 genes that would represent between 72% and 84% of their respective genomes. The genome comparison between M31^T and M8 [51], at the time of its publication, was the closest intraspecific evaluation ever undertaken, and already showed such coexisting strains were significantly different because of unique genes, genomic islands and the plasmid content. Unique genes within each single strain varied between 148 and 556, which represented between 5% and 17% of each individual genome (Figure 4 and Supplementary Spreadsheet). Intergeneric comparisons within the same *Salinibacteraceae* family (i.e. between the strains of the genera *Salinibacter, Salinivenus* and *Salisaeta*) showed that the core genomes ranged between 1,857 and 1,387 (Figures 4E and 4F) that would represent >41% of the gene content for each single genome.

Approximately 4,800 of the pan-genome genes (i.e. 53%) were strain-specific, and about 2,300 of them were specific to the most distantly related genomes of *Sat. longa* (813 genes; GCA_000419585.1) and *Rho. marinus* (1,509 genes; CP001807.1). In all cases, the number of non-annotated ORFs or hypothetical proteins was always >50% of the strain-unique genes, and those annotated were mostly related to either mobile elements (such as transposases), viruses or their infection mechanisms (such as integrases or CRISPR associated proteins), plasmids or some different sulfo-, methyl- or glycotransferases (Supplementary Spreadsheet). These types of genes have been observed to be abundant in genomic islands, and related to distinct salinity concentration performances [51]. However, no especially conspicuous metabolism discriminating the different phylotypes based on the gene composition could be observed, and the differences in gene repertoire and sequence divergences between orthologous genes could have been related only to distinct performances of similar metabolisms. For instance, the major phenotypic differences observed between the Argentinian isolates and *Sal. ruber* were mostly associated with their tolerances to salt, temperature or pH (see below), and could not be attributed to distinct gene content, but rather to distinct expression of orthologous genes.

The *Salinibacteraceae* family genome collection encoded for a set of 357 specific genes, 39 (11%) of which had been annotated as hypothetical proteins in M31^T (Figure 5; Supplementary Spreadsheet). The remaining family-specific repertoire, which was not present in *Rho. marinus*, also encoded for certain central metabolisms, such as amino acid metabolism (e.g. glycine, cysteine, lysine) that could also be related to the osmotic stress response [59]. Other genes were related to carbohydrate metabolism (such as some genes from the pentose phosphate pathway, or from the serine-glyoxylate cycle), or DNA repair systems. Some genes were also found to be related to carotenoid biosynthesis, which were possibly responsible for the pigmentation of their colonies [34, 39], the cobalt-cadmium-zinc resistance proteins, or an arsenic-driving pump related to arsenic resistance. Actually, arsenic is known to be present in hypersaline environments located in high altitude Argentinian lakes [17, 29, 30, 56], and

microorganisms living in these extreme environments encode genes related to inorganic arsenite (AsO_2H) and arsenate (AsO_4^{3-}) detoxification of the *ars* operon [47].

The *Salinibacter – Salinivenus* lineage shared a set of 349 genes, 90 of which (26%) where hypothetical proteins (Figure 5; Supplementary Spreadsheet). The remaining genera-specific repertoire, not present in *Rho. marinus* or *Sat. longa*, also encoded for central metabolism genes, such as those involved in amino acid synthesis, DNA repair or carotenoid biosynthesis. Genes for potassium homeostasis and the multi-subunit cation (Na^+/H^+) antiporter complex were also found. Some components of the latter, but not the complete operons, were also detected in *Sat. longa*. In this study, and in the whole family specific genes, different components were also found for choline and betaine uptake and betaine biosynthesis related to the osmotic stress responses, as well as some genes related to the cobalt-cadmium-zinc resistance proteins. Such findings would agree with the lower salinity tolerance of *Sat. longa* and its unclear origin (Red Sea or Dead Sea [68]).

When focusing only on *Salinibacter* genus-specific genes a core set of 317 CDS was found, with 103 of them (33%) being hypothetical proteins (Figure 5; Supplementary Spreadsheet). The remaining species-specific repertoire encoded mostly for unclear functions that could not be annotated using the SEED database. From the annotated genes, the flagellum synthesis components that appear in accordance with the detected motility of *Sal. ruber* and the new Argentinian isolates, but not the *Salinivenus* species [36] or probably *Salisaeta*, did not have the motility phenotype assessed [68]. However, there were some genes exclusive to *Sal. ruber* that were related to iron acquisition by hemin transport systems.

The two different phylotypes of *Sal. ruber* EHB-1 and EHB-2 encoded a small set of 33 (20 hypothetical ORFs; 61%) and 79 (54 hypothetical ORFs; 68%) phylotype-specific genes, respectively. The remaining phylotype-specific repertoire in both cases encoded for either mobile elements (such as transposases), or virus association (such as recombinases). As indicated below, *Sal. ruber* was also the single group detected encoding for a halorhodopsin [51] and a unique lineage of sensory rhodopsin type I (Supplementary Table S3, and Supplementary Spreadsheet).

Each single species accounted for a unique species-specific gene repertoire (Figure 4; Supplementary Spreadsheet) consisting of 101 (63 hypothetical ORFs; 63%) genes for *Sal. ruber*, 138 (69 hypothetical ORFs; 50%) for the Argentinian strains, and 353 (150 hypothetical ORFs; 43%) for *Salinivenus* (considering both *Salinivenus* species as one single species). The annotated species-specific repertoire did not encode for any conspicuous functional genes or pathways. Only an apparent high number of orthologous genes related to histidine kinases and transferases (especially glycosyltransferases) was remarkable for each individual species. As mentioned above, these results indicated that the divergence between the different species

studied was mainly due to the evolution of orthologous genes that will ultimately produce different metabolic performances. Gene acquisition and/or gene loss did not seem to be important as responsible for the distinct ecological capabilities. Consistent with these findings, the pan-genome of the genus was predominately comprised of hypothetical and mobile genes (e.g. >70% of the total genes when the core genes were removed from the analysis).

Genome synteny

The synteny of the gene order along the chromosomes decreased with the evolutionary distance (Figure 6), which was similar to that observed for the genomic architecture of *Campylobacter* [22]. When taking M31^T as a reference, a sharp decrease in the percentages of the conservation order with the evolutionary distance could be observed (Figure 6; Supplementary Table S4). EHB-1 phylotypes conserved 87.3% of the gene order, as already observed [51], and this was similar between phylotypes 84%. The synteny decreased to 14.7% of the gene order conservation with the Argentinian isolates, but it was already possible to observe some large genome rearrangements showing at least two relevant inversions (Figure 6). Synteny between both genera *Salinibacter* and *Salinivenus* still showed a conserved gene order (3.5%), but it disappeared completely when taking into account the least related genus *Salisaeta* (only 0.53%).

***Sal. ruber* EHB-2 strain ST67 closed genome**

The EHB-2 strain ST67 genome was fully closed and showed two plasmids: ST67-pSR1 with a size of 178 kb encoding 128 CDS, and ST67-pSR2 with a size of 91 kb encoding 69 CDS with GC% values of 60.1% and 60.4%, respectively, which were lower than the chromosome (66%). *Sal. ruber* M8 and M31^T had 4 and 1 plasmids, respectively, with a GC% ranging between 59 and 63%, also with lower GC% [51]. The sequencing depth of the chromosome was 30X and close to 60X for the plasmids, suggesting that the plasmids were present in two copies each. In plasmid ST67-pSR1, 28% of the proteins were hypothetical, whereas they were 46% in plasmid ST67-pSR2. Both plasmids encoded for the *parA* gene involved in the partition and division of the plasmid, as well as for a replication initiation protein and one recombinase. In addition, both plasmids shared 36 orthologous genes: 6 annotated as transposases, 1 transcriptional regulator, 8 hypothetical proteins, 1 recombinase and 5 transporter membrane proteins. Plasmid ST67-pSR1 shared between 12 and 37 orthologous genes with the M8 plasmids and 26 with the single plasmid of M31^T. Plasmid ST67-pSR2 shared between 8 and 29 orthologous genes with the M8 plasmids and 21 with the plasmid of M31^T.

In addition, ST67 had three genomic islands not present in any of the studied genomes (Supplementary Tables S5, S6, and Supplementary Figure S4), two of them encoding for >60 CDS, and the third only for 8 CDS. The gene annotation of the major categories showed the presence of genes related to mobile elements, such as phage genes and transposases. ST67-HVR1 encoded for glycosyltransferase and sulfotransferase genes, which was similar to that

occurring in M8 that discriminated and conferred a different phenotype from M31^T [51]. Conspicuously, ST67-HVR2 encoded for a type VI secretion system (T6SS) that had often been previously observed in *Proteobacteria* and was hypothesized to originate from lateral gene transfer [10]. The 11 genes encoding for the T6SS had their closest syntenic structure in the *Rhodothermaceae* strain RA (Supplementary Figure S5 [23]). This contact-dependent armament against other bacteria or eukaryotes has also been found in neighboring phylum members of gut *Bacteroidetes* with three genetic architectures, two of which appear encoded as conjugative elements [12, 15]. However, in ST67, the system observed was chromosomally encoded and the closest similar relative T6SS annotated was in strain RA, which was a *Rhodothermaeota* strain isolated from a Malaysian hot spring very loosely affiliated with the *Salinibacteraceae* [23]. However, only these two genomes (*Sal. ruber* ST67 and *Rhodothermaceae* strain RA) exhibited this T6SS within the *Rhodothermaeota*, although the relatively high similarity between both may respond to a phylum-specific system. Most of the T6SS effectors seem to target Gram-negative bacteria, but this may be because they have been studied and found mostly in *Proteobacteria* [10, 19]. However, the recent finding of this system in *Bacteroidetes*, and especially among co-occurring gut *Bacteroidales*, indicates that it would be an effective antagonistic system against common competitors or predators [12]. Similarly, such a system in ST67 would help in antagonizing other members of the same brine community that could compete or even predate it, conferring an ecological advantage towards other coexisting *Salinibacter* members. To our knowledge this is the first report of a T6SS in *Rhodothermaeota*.

Special features

Relevant features were searched for in the genomes, and one of the most conspicuous genes detected in the first genome sequencing was the presence of different types of rhodopsins [39], especially xanthorhodopsin [5]. As we have already shown, not all *Sal. ruber* isolates contained the same rhodopsin dosage [52]. In this current study, it was observed (Table S3 and Figure S6) that only the very characteristic xanthorhodopsin [5] was universally present in the complete family with a genealogy in accordance with that of the housekeeping genes. However, halorhodopsins still seemed to be exclusive to some members of *Sal. ruber* EHB-1 [51, 52], whereas sensory rhodopsins were exclusive to the genus *Salinibacter* (Table S3 and Figure S6). One divergent lineage of the sensory rhodopsin I was exclusive to this EHB-1 phylotype. The two isolates of the second phylotype of *Sal. ruber* differed in their rhodopsin gene content. ST67 only encoded for xanthorhodopsin and one sensory rhodopsin type I, and SP273, in addition to xanthorhodopsin, encoded for three divergent paralogous sensory type I rhodopsins. The three Argentinian strains encoded for one sensory rhodopsin and one xanthorhodopsin. Finally, *Slv. iranica* (CB7^T), *Slv. lutea* (DGO^T) and *Sat. longa* (S4-4^T) only encoded for a single xanthorhodopsin. No rhodopsin was found in *Rho. marinus* (DSM 4252^T). Nevertheless, only xanthorhodopsin was the single orthologous common to all *Salinibacteraceae*, whose phylogeny mirrored the host housekeeping genes indicating that it has been maintained from

the last common ancestors of the strains. The remaining rhodopsins were distinctly present in *Salinibacter* strains, and their genealogies also mirrored the genome genealogies, indicating they could be horizontally acquired in distinct speciation stages of the members of the genus. [49].

A previous phylogenomic study identified the presence of 40 genes in the genome of *Sal. ruber* M8 that were likely to be involved in lateral gene transfer events between Archaea and Bacteria [50]. In a previous study [49], we screened 92 *Sal. ruber* EHB-1 strains and detected the presence of most of these putative LGT genes (from 25 to 40) in each strain. A principal component analysis based on presence/absence of these genes showed a distribution of strains related to their origin of isolation, which could have a similar history of gene transfer from Archaea carried in the genomes of geographically related strains [49]. In this current study, the presence of these genes was analyzed (Supplementary Table S7 and Supplementary Figure S7) and it was found that the presence/absence pattern strongly supported the three genera classification proposed here.

Finally, the presence of the CRISPR-Cas system was investigated since it was described originally in the halophilic archaeon *Haloferax mediterranei* [38] that has been recognized as an extremely relevant immunity system for prokaryotes [35]. All genomes studied here, except those of *Sal. ruber* EHB-1 strains M31^T, M8 and EHB-2 strain ST67, encoded for CRISPR-Cas systems (Table 3 and Supplementary Figure S3). The three strains of *Sal. altiplanensis* and the two members of *Salinivenus, Sat. longa* and *Rho. marinus*, had a homologous CRISPR-Cas system that could be classified as Type I-E [35] from the Cas1 genealogy (Supplementary Figure S3). The Argentinian strains, both *Salinivenus* species, *Salisaeta* and *Rhodothermus*, encoded for the Type I-E system, *Sal. ruber* SP273 for a Type I-B system, and *Slv. iranica* encoded in addition a second operon related to a Type I-C system. Actually, and very conspicuously, this genealogy mirrored the expectation for a vertical heritage, which would be the only parsimonious explanation given the large distance between their geographic origins (South American Altiplano for *Sal. altiplanensis*, Iranian salt lake for *Salinivenus*, and Malaysia for *Rho. marinus*), and would presumably act as a barrier to gene exchange. Conspicuously, *Slv. iranica*, which coexisted with *Slv. lutea* [36], encoded for a unique Type 1-C system. On the other hand, the single isolate of *Sal. ruber* encoding for CRISPR-Cas was SP273 (Supplementary Figure S3), and it encoded for a totally different operon, not found in any other *Salinibacteraceae* genome, which, according to the Cas1 genealogy, could be classified as Type I-B [35].

Altogether, at least 28 different spacer types could be detected and each organism exhibited between one to four different spacer regions. None of these spacers matched with viruses isolated from *Sal. ruber* [69], nor mapped to the different metaviromes currently generated in the laboratory (data not shown). However, the involvement of CRISPR-spacers in *S. ruber*

resistance to virus has been shown recently [69]. Interestingly, they were all different, except for two shared by the Argentinian strains LL19 and AN4 (Table 3). Both strains had been isolated from the Argentinian Altiplano, but from distinct hydrographical basins separated only by approximately 60 km. The fact that they shared two identical spacers may be evidence of having been infected by similar viruses. Thus, either both strains shared an ecosystem or the viruses that infected them. The dispersal mechanisms of halophiles in such high mountain hypersaline lakes are unknown, but microorganisms could have migrated along the short distance either through aerial dispersion by wind flow [26] or simply by the colonization of suitable hypersaline animal environments, such as bird salt nostrils [11], or by mechanical transport attached to feathers of common migratory inhabitants, such as flamingos [16].

Taxonomic characterization of the new isolates

Since the genomic analyses showed that the new Argentinian isolates could be classified as members of a new species of *Salinibacter,* strain AN15^T was designated as the type strain proposed as *Salinibacter altiplanensis* sp. nov. (Table 4). The phenotypic characterization was carried out in accordance with the known metabolic profiles [3, 36]. Table 1 and Supplementary Table S8 show the phenotypic profiles of all representative type strains of the genus *Salinibacter* and *Salinivenus*. From the main characteristics, the Argentinian strains could be considered as members of the genus *Salinibacter*, and the most prominent diagnostic characters were that they were extremely halophilic bacteria with optimum growth at 25% salinity (5% higher than *Sal. ruber*), Gram-negative forming red colonies on SW agar medium approximately 0.8 mm in diameter after three weeks growth. The cells were curved rods, motile, with a smaller cell size (3.2 – 4 µm in length), lower growth temperature optimum (30 ºC) and pH optimum (7.5) than the neighboring species *Sal. ruber*. Their genome size ranged from 3.58 – 3.71 Mb with a G+C% between 64.2% and 64.4%. Phenotypically the isolates were catalase and oxidase, lysine decarboxylase, ornithine decarboxylase and DNA hydrolysis positive; but negative for hydrolysis of Tween 20, casein hydrolysis, arginine hydrolase, indole production, Voges-Proskauer test, methyl red test, production of H_2S, or gas formation with nitrate. D-fructose, D-glucose and sucrose were used by the three strains; D-galactose, L-lysine, D-mannitol, L-proline, L-tryptophan, maltose and raffinose were variable; whereas D-ribose, lactose, L-alanine, L-asparagine, L-aspartate, L-cysteine, L- methionine, L-phenylalanine and L-tyrosine were not used by any of the strains. The fatty acid profiles (Supplementary Table S9) were not particularly discriminative, with $C_{15:0}$ iso, $C_{15:0}$ anteiso, $C_{16:0}$, $C_{17:0}$ iso 3-OH and $C_{18:1}$ ω7c being the most relevant. Only $C_{15:0}$ iso appeared slightly below the percentages determined for *Sal. ruber* strains, but the overall profiles were very similar among all *Salinibacter* strains, and in accordance with the genus description [36]. In addition, the biomass showed a maximum absorption of pigments at 481 nm. Finally, and in order to identify the members of this new species *Sal. altiplanensis* directly using phylogenetic probes, a specific EHB-130 probe 5'-CTTTTGGGCAGGTTGTCT-3' (starting the target complementary sequence in the *E. coli* position 130) was designed and optimized. For fluorescence *in situ* hybridization, the probe was

optimally hybridized with a formamide concentration of 35%, and Figure 7 shows the probe specificity of the new species *Sal. altiplanensis*. AN15^T was designated as the type strain and it was deposited in two international strain collections (=CECT 9105^T = IBRC-M 11031^T). This type strain was isolated from the Antofalla salt lake, located in the Argentinian Altiplano at an altitude of 4,000 m (Argentina). The formal proposal of the new species is provided in the protologue (Table 4), which represents information extracted from Digital Protologue Taxonumber TA00140.

Acknowledgements

This study was funded by the Spanish Ministry of Economy projects CGL2012-39627-C03-03 and CLG2015_66686-C3-1-P (to RRM), CLG2015_66686-C3-3-P (to JA), which were also both supported with European Regional Development Fund (FEDER) funds. RA was financed by the Max Planck Society. KTK's research was supported, in part, by the U.S. National Science Foundation (Award No. 1241046). TVP acknowledges a pre-doctoral fellowship (Nr BES-2013-064420) from the Ministerio de Economía y Competitividad of the Spanish Government. The authors also acknowledge the economic support and unconditional scientific interest of Lipotrue SL and Deep Blue Sea SL.

References

[1] Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

[2] Antón, J., Lucio, M., Peña, A., Cifuentes, A., Brito-Echeverría, J., Moritz, F., Tziotis, D., López, C., Urdiain, M., Schmitt-Kopplin, P., Rosselló-Móra, R. (2013) High metabolomic microdiversity within co-occurring isolates of the extremely halophilic bacterium *Salinibacter ruber*. PLoS ONE 8(5): e64701.

[3] Antón, J., Oren, A., Benlloch, S., Rodríguez-Valera, F., Amann, R., Rosselló-Móra, R. (2002) *Salinibacter ruber* gen. nov., sp. nov., a novel extremely halophilic member of the Bacteria from saltern crystallizer ponds. Int. J. Sys. Evol. Microbiol. 52**,** 485-491.

[4] Antón, J., Rosselló-Móra, R., Rodriguez-Valera, F., Amann, R. (2000) Extremely halophilic bacteria in crystallizer ponds from solar salterns. Appl. Environ. Microbiol. 66, 3052-3057.

[5] Balashov, S.P., Imasheva, E.S., Boichenko, V.A., Antón, J., Wang, J.M., Lanyi, J.K. (2005) Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. Science 309, 2061-2064.

[6] Bankevich, A., Nurk, S., Antipov, D., Gurevich A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko S.I., Pham, S., Prjibelski, A.D., Pyshkin A.V., Sirotkin A.V., Vyahhi, N., Tesler, G., Alekseyev M.A., Pevzner P.A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455-477.

[7] Barrow, G., Feltham, R. (Eds.). (1993). Cowan and Steel's Manual for the Identification of Medical Bacteria. Cambridge: Cambridge University Press.

[8] Besemer, J., Lomsadze, A., Borodovsky, M. (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 20, 2607-2618.

[9] Boetzer, M., Pirovano, W., (2014). SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics 15, 211.

[10] Boyer, F., Fichant, G., Berthold, J., Vandrenbrouck, Y., Attree, I. (2009) Dissecting the bacterial type VI secretion system by a genome wide *in silico* analysis: what can be learned from available microbial genomic resources? BCM Genomics. 10, 104.

[11] Brito-Echeverría, J., López-López, A., Yarza, P., Antón, J., Rosselló-Móra, R. (2009) Occurrence of *Halococcus* spp. in the nostrils salt glands of the seabird *Calonectris diomedea*. Extremophiles. 13, 557-565.

[12] Chatzidaki-Livanis, M., Geva-Zartosky, N., Comstock, L.E. (2016) *Bacteroides fragilis* type VI secretion systems use novel effector and immunity proteins to antagonize human gut Bacteroidales species. Proc. Natl. Acad. Sci. USA. 113, 3627-3632.

[13] Conesa, A., Götz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics. 619832.

[14] Cox, M.P., Peterson, D.A., Biggs, P.J. (2010) SolexaQA: at-a-glance quality assessment of Illumina second generation sequencing data. BMC Bioinformatics 11, 485.

[15] Coyne, M.J., Roelofs, K.G., Comstock, L.E. (2016) Type VI secretion systems of human gut Bacteroidales segregate into three genetic architectures, two of which are contained on mobile genetic elements. BMC Genomics. 17, 58.

[16] Derlindati, E., Romano, M.C., Cruz, N.N., Barisón, C., Arengo, F., Barberis, I.M. (2014) Seasonal activity patterns and abundance of Andean flamingo (*Phoenicoparrus andinus*) at two contrasting wetlands in Argentina. Ornitol. Neotrop. 25, 317-331.

[17] Dib, J., Motok, J., Zenoff, V.F., Ordoñez, O., Farías, M.E. (2008) Occurrence of resistance to antibiotics, UV.B, and arsenic in bacteria isolated from extreme environments in high-altitude (above 4400 m) Andean wetlands. Curr. Microbiol. 56, 510-517.

[18] Di Meglio, L., Santos F., Gomariz, M., Almansa, C., López, C., Antón, J., Nercessian, D. (2016) Seasonal dynamics of extremely halophilic microbial communities in three Argentinian salterns. FEMS Microbiol. Ecol. 92, fiw184.

[19] Durand, E., Cambillau, C., Cascales, E., Journet, L. (2014) VgrG, Tae, Tle, and beyond: the versatile arsenal of Type VI secretion effectors. Trends Microbiol. 22, 498-507.

[20] Dussault, H.P. (1955) An improved technique for staining red halophilic bacteria. J Bacteriol. 70, 484-485.

[21] Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32, 1792–1797.

[22] Fouts, D.E., Mongodin, E.F., Mandrell, R.E., Miller, W.G., Rasko, D.A., Ravel, J., Brinkac, L.M., DeBoy, R.T., Parker, C.T., Daugherty, S.C., Dodson, R.J., Durkin, A.S., Madupu, R., Sullivan, S.A., Shetty, J.U., Ayodeji, M.A., Shvartsbeyn, A., Chatz, M.C., Badger, J.H., Fraser, C.M., Nelson, K.E. (2005) Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species. PLoS Biol 3, e15.

[23] Goh, K.M., Chan, K.G., Lim, S.W., Liew, K.J., Chan, C.S., Shamsir, M.S., Ee, R., Adrian, T.G. (2016) Genome analysis of a new *Rhodothermaceae* strain isolated from a hot spring. Front. Microbiol. 7, 1109.

[24] Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57, 81-91.

[25] Grissa, I., Vergnuad, G., Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 35, W52-W57.

[26] Kellogg, C.A., Griffin, D.W. (2006) Aerobiology and the global transport of desert dust. Trends Ecol. Evol. 21, 638-644.

[27] Konstantinidis, K., Tiedje, J.M. (2007) Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Curr. Opin. Microbiol. 10, 504–509.

[28] Konstantinidis, K.T., Tiedje, J.M. (2005). Towards a genome-based taxonomy for prokaryotes. J. Bacteriol. 187, 6258–6264.

[29] Kurth, D., Amadio, A., Ordoñez, O.F., Albarracín, V.H., Gärtner, W., Farías, M.E. (2016). Arsenic metabolism in high altitude modern stromatolites revealed by metagenomic analysis. Scientific Reports. 7, 1024.

[30] Kurth, D., Belfiore, C., Gorriti, M.F., Cortez, N., Farias, M.E., Albarracín, V.H. (2015) Genomic and proteomic evidences unravel the UV-resistome of the poly-extremophile *Acinetobacter* sp. Ver3. Front. Microbiol. 6, 308.

[31] Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. Genome Biol. 5, R12.

[32] Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Liss, T., Lussmann, T., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vildig, A., Lenke, M., Ludwig, T., Bode, A., Schleifer, K.H. (2004) ARB: a software environment for sequence data. Nucleic Acids Res. 32, 1363-1371.

[33] Luo, C., Tsementzi, D., Kyripides, N.C., Konstantinidis, K.T. (2012) Individual genome assembly from complex community short-read metagenomic datasets. ISMEJ. 6, 898-901.

[34] Lutnaes, B.F., Strand, A., Pétursdóttir, S.K., Liaaen-Jensen, S. (2004) Carotenids of thermophilic bacteria-*Rhodothermus marinus* from submarine Icelandic hot springs. Biochem. Syst. Ecol. 32, 455-468.

[35] Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.M., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., Hovarth, P., Moineau, S., Mojica, F.J.M., Terns, R.M., Terns, M.P., White, M.F., Yakunin, A.F., Garrett, R.A., van der Oost, J., Backofen, R., Koonin, E.V. (2015) An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. 13, 722-736.

[36] Makhdoumi-Kakhki, A., Amoozegar, M.A., Ventosa, A. (2012) *Salinibacter iranicus* sp. nov. and *Salinibacter luteus* sp. nov., isolated from a salt lake, and emended descriptions of the genus *Salinibacter* and of *Salinibacter ruber*. Int. J. Syst. Evol. Microbiol. 62, 1521-1527.

[37] Mira, A., Martín-Cuadrado, A,B., D'Auria, G., Rodríguez-Valera, F. (2010) The bacterial pan-genome: a new paradigm in microbiology. Int. Microbiol. 13, 45-57.

[38] Mojica, F.J., Rodriguez-Varela, F. (2016) The discovery of CRISPR in archaea and bacteria. The FEBS Journal. 283, 3162-3169.

[39] Mongodin, E.F., Nelson K.E., Daugherty, S., DeBoy, R.T., Wister, J., Khouri, H., Weidman, J., Walsh, D.A., Papke, R.T., Sanchez-Perez, G., Sharma, A.K., Nesbø, MacLeod, D., Bapteste, E., Doolittle, F.F., Charlebois, R.L., Legaut, B., Rodríguez-Valera, F. (2005) The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. Proc. Natl. Acad. Sci. USA 102, 18147-18152.

[40] Mora-Ruiz, M.R., Cifuentes, A., Font-Verdera, F., Pérez-Fernández, C., Farías, M.E., González, B., Orfina, A., Rosselló-Móra, R. (2017) Biogeographical patterns of bacterial and archaeal communities of distant hypersaline environments. Syst. Appl. Microbiol. IN PRESS

[41] Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A., Kanehisa, M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182–W185.

[42] Munoz, R., Rosselló-Móra, R., Amann, R. (2016) Revised phylogeny of *Bacteroidetes* and proposal of sixteen new taxa and two new combinations including *Rhodothermaeota* phyl. nov. Syst. Appl. Microbiol. 39, 281-296.

[43] Mutlu, M.B., Martínez-García, M., Santos, F., Peña, A., Guven, K., Antón, J. (2008) Prokaryotic diversity in Tuz lake, a hypersaline environment in inland Turkey. FEMS Microbiol. Ecol. 65, 474-483.

[44] Nolan, M., Tindall, B.J., Ponrenke, H., Lapidus, A., Copeland, A., Glavina, T., Lucas, S., Chen, F. (2009) Complete genome sequence of *Rhodothermus marinus* type strain (R-10^T). Stand. Genomic Sci. 1, 283-290.

[45] Oren, A. (2013) *Salinibacter*: an extremely halophilic bacterium with archaeal properties. FEMS Microbiol. Lett. 342, 1-9.

[46] Oren, A., Ventosa, A. (2000) International Committee on Systematic Bacteriology – Subcommittee on the taxonomy of *Halobacteriaceae*. Minutes of the meetings, 16 August 1999, Sydney, Australia. Int. J. Syst. Evol. Microbiol. 50, 1405-1407.

[47] Páez-Espino, D., Tamames, J., de Lorenzo, V., Cánovas, D. (2009) Microbial responses to environmental arsenic. Biometals. 22, 117-130.

[48] Peng, Y., Leung, H.C., Yiu, S.M., Chin, F.Y. (2012) IDBA-UD: a *de novo* assembler for single-cell and metagenomics sequencing data with highly uneven depth. Bioinformatics. 28, 1420–1428.

[49] Peña, A., Gomariz, M., Lucio, M., González-Torres, P., Huertas-Cepa, J., Martínez-García, M., Santos, F., Schmitt-Kopplin, P., Gabaldón, T., Rosselló-Móra, R., Antón, J. (2014) *Salinibacter ruber*: the never ending microdiversity? In: Papke, T., Oren, A., Ventosa, A. (eds). Genetics and Genomics of Halophiles. Caister Academic Press: Norfolk, UK, pp 37-53.

[50] Peña, A., Teeling, H., Huerta-Cepas, J., Santos, F., Meseguer, I., Lucio, M., Schmitt-Kopplin, P., Dopazo, J., Rosselló-Móra, R., Schüler, M., Glöckner, F.O., Amann, R., Gabaldón T., Antón, J. (2011) From genomics to microevolution and ecology: the case of *Salinibacter ruber. In:* Halophiles and Hypersaline Environments (Ventosa, A., et al. eds) Springer Verlag (Berlin – Heilderberg). DOI 10.1007/978-3-642-20198-1_5. ISBN: 978-3-642-20197-4. pp. 109-122.

[51] Peña, A., Keeling, H., Huertas-Cepas, J., Santos, F., Yarza, P., Brito-Echeverría, J., Lucio, M., Schmitt-Kopplin, P., Meseguer, I., Schenowitz, C., Dossat, C., Barbé, V., Dopazo, J., Rosselló-Móra, R., Schüler, M., Glöckner, F.O., Amann, R., Gabaldón, T., Antón, J. (2010) Fine-scale evolution: genomic, phenotypic and ecological differentiation in two coexisting *Salinibacter ruber* strains. ISMEJ. 4, 882-895.

[52] Peña, A., Valens, M., Santos, F., Buczolits, S., Antón, J., Kämpfer, P., Busse, H.J., Amann, R., Rosselló-Mora, R. (2005) Intraspecific comparative analysis of the species *Salinibacter ruber.* Extremophiles. 9**,** 151-161.

[53] Pernthaler, A., Pernthaler, J., Amann, R. (2002) Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. Appl. Environ. Microbiol. 68, 3094–3101.

[54] Podell, S., Ugalde, J.A., Narasingarao, P., Banfield, J., Heidelberg, K.B., Allen, E.E. (2013) Assembly-driven community genomics of a hypersaline microbial ecosystem. PLoS ONE 8:e61692.

[55] Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35, 7188-7196.

[56] Rascovan, N., Maldonado, J., Vazquez, M. P., Farías, E. (2016) Metagenomic study of red biofilms from Diamante Lake reveals ancient arsenic bioenergetics in haloarchaea. ISME J. 10, 299-309.

[57] Richter, M., Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc. Nat. Acad. Sci. USA. 106, 19126-19131.

[58] Rissman, A.I., Mau, B., Biehl, B.S., Darling, A.E., Glasner, J.D., Perna, N.T. (2009) Reordering contigs of draft genomes using the Mauve Aligner. Bioinformatics. 25, 2071–2073.

[59] Roberts, M.F. (2005) Organic compatible solutes of halotolerant and halophilic microorganisms. Saline Syst. 1, 5.

[60] Rodriguez-R, L.M., Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ. Preprints. 4, e1900v1.

[61] Rodriguez-Valera, F., Ventosa, A., Juez, G., Imhoff, J.F. (1985) Variation of environmental features and microbial populations with salt concentrations in a multi-ponds saltern. Microbial. Ecol. 11, 107-115.

[62] Rosselló-Móra, R., Amann, R. (2015) Past and future species definitions for *Bacteria* and *Archaea*. Syst. Appl. Microbiol. 38, 209-216.

[63] Rosselló-Móra, R., Lucio, M., Peña, A., Brito-Echeverría, J., López-López, A., Valens-Vadell, M., Frommberger, M., Antón, J., Schmitt-Kopplin, P. (2008) Metabolic evidence for biogeographic isolation of the extremophilic bacterium *Salinibacter ruber*. ISMEJ. 2, 242-253.

[64] Saitou, N., Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic tress. Mol. Biol. Evol. 4, 406-425.

[65] Soria-Carrasco, V., Valens-Vadell, M., Peña, A., Antón, P., Amann, R., Castresana, J., Rosselló-Mora, R. (2007) Phylogenetic position of *Salinibacter ruber* based on concatenated protein alignments. Syst. Appl. Microbiol. 30, 171-179.

[66] Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22, 2688-2690.

[67] Urdiain, M., López-López, A., Gonzalo, C., Busse, J., Langer, S., Kämpfer, P., Rosselló-Móra, R. (2008) Reclassification of *Rhodobium marinum* and *Rhodobium pfennigii* as *Afifella marina* gen. nov. comb. nov. and *Afifella pfennigii* comb. nov, a new genus of photoheterotrophic *Alphaproteobacteria* and emended descriptions of *Rhodobium, Rhodobium orientis* and *Rhodobium gokarnense*. Syst. Appl. Microbiol. 31, 339-351.

[68] Vaisman, N., Oren A. (2009) *Salisaeta longa* gen. nov., sp. nov., a red, halophilic member of the *Bacteroidetes*. Int. J. Syst. Evol. Microbiol. 59, 2571-2574.

[69] Villamor, J., Ramos-Barbero, M.D., González-Torres, P., Gabaldón, T., Rosselló-Móra, R., Meseguer, I., Martínez-García, M., Santos, F., Anton, J. (2017) Characterization of ecologically diverse viruses infecting co-occurring strains of cosmopolitan hyperhalophilic *Bacteroidetes.* ISMEJ. Ahead of publication doi:10.1038/ismej.2017.175

[70] Viver, T., Cifuentes, A., Díaz, S., Rodríguez-Valdecantos, G., González, B., Antón, J., Rosselló-Móra, R. (2015) Diversity of extremely halophilic cultivable prokaryotes in Mediterranean, Atlantic and Pacific solar salterns: evidence that unexplored sites constitute sources of cultivable novelty. Syst. Appl. Microbiol. 38, 266-275.

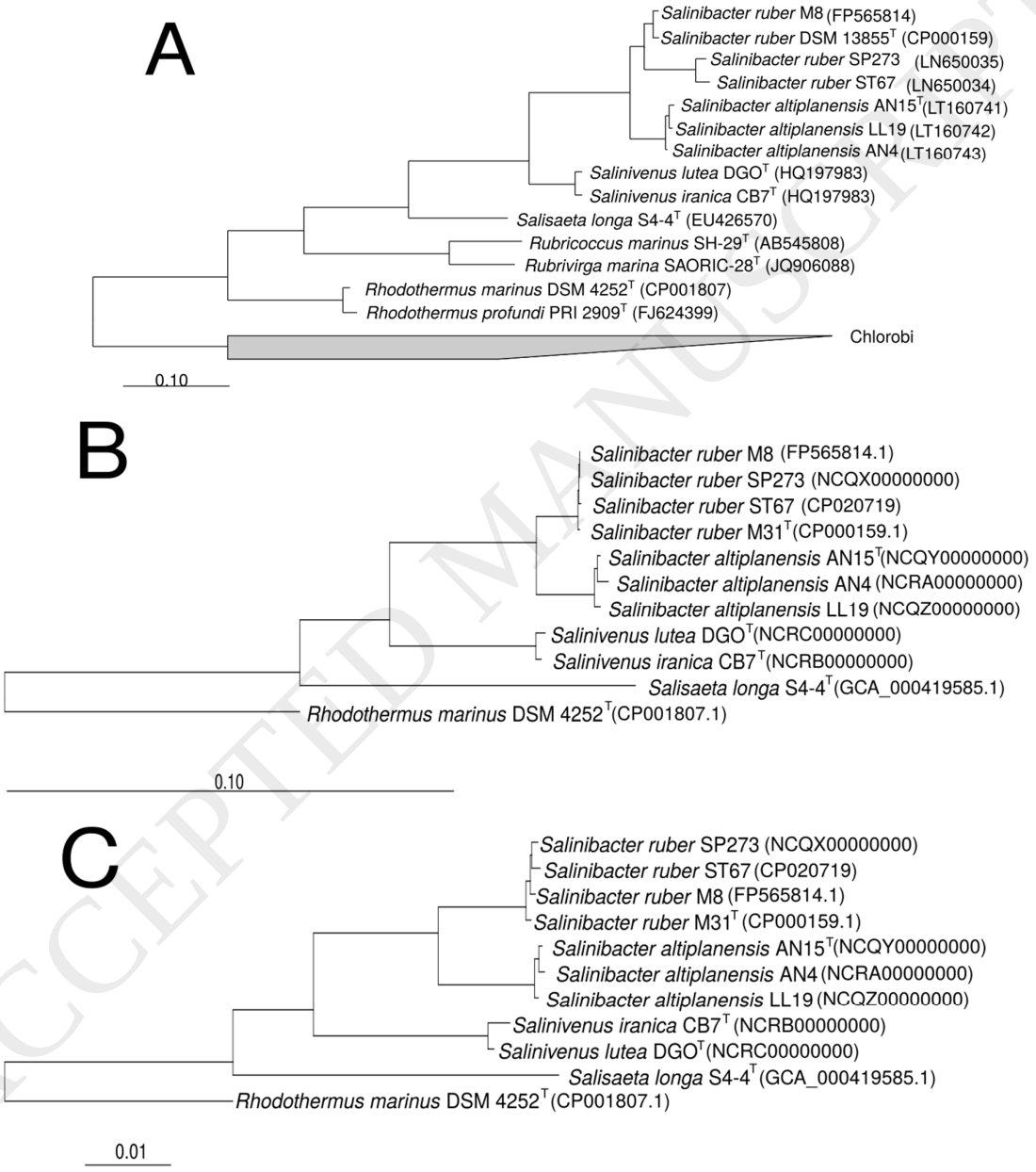
[71] Wickham, H., Chang, W. (2016) Package “ggplot2”.

[72] Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.H., Whitman, W.B., Euzéby, J., Amann, R., Rosselló-Móra, R. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nature Rev. Microbiol. 12, 635-645.

[73] Zerbino, D.R., Birnew E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18, 821-829.

[74] Zhong, Z.P., Liu, Y., Miao, L.L., Wang, F., Chu, L.M., Wang, J.L., Liu, Z.P. (2016) Prokaryotic community structure driven by salinity and ionic concentrations in plateau lakes of the Tibetan Plateau. Appl. Environ. Microbiol. 82, 1846-1858.

Figure 1: Phylogenetic tree reconstructions based on a neighbor-joining calculation for the (A) 16S rRNA genes; (B) 23S rRNA genes; (C) MLSA concatenated genes; and (D) the core-genes of the *Salinibacteraceae* genomes and *Rho. marinus*. The 23S rRNA genes, MLSA genes and core-genes were extracted from the genome sequence (acc. no. genome in brackets).



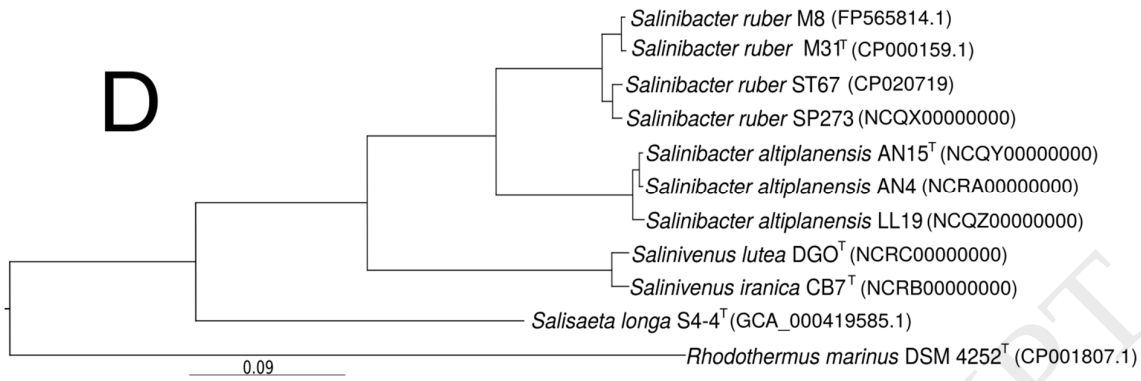


Figure 2: Correlation between the ANI and AAI values of pairwise compared genomes with the 16S rRNA gene sequence identities and with the percentages of shared genomic DNAs. (A) ANI vs 16S rRNA gene identities; (B) AAI vs 16S rRNA gene identities; (C) number of shared genes between each pairwise genome comparison vs ANI; (D) number of shared proteins between each pairwise genome comparison vs AAI. The species boundary was considered to be the ANI range between 93% and 96% (Richter and Rosselló-Móra, 2009; [57]), and the genus boundary for AAI the range between 67% and 73% (Konstantinidis and Tiedje, 2008)

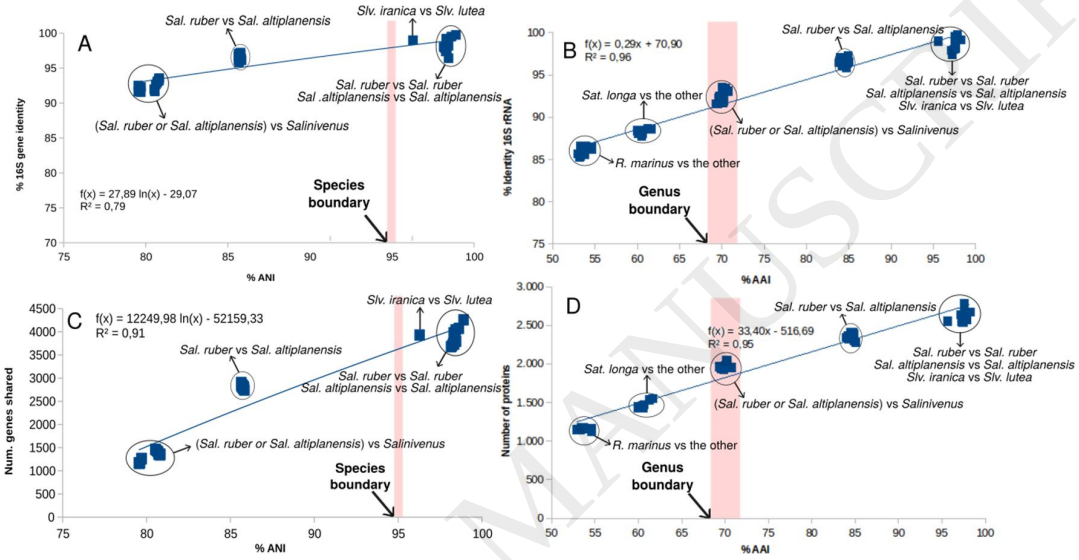


Figure 3: Core- and pan-genome configuration with the increase of genomes in the study. The pan-genome considered here accounts for all available genomes in the *Rhodothermaceae* phylum.

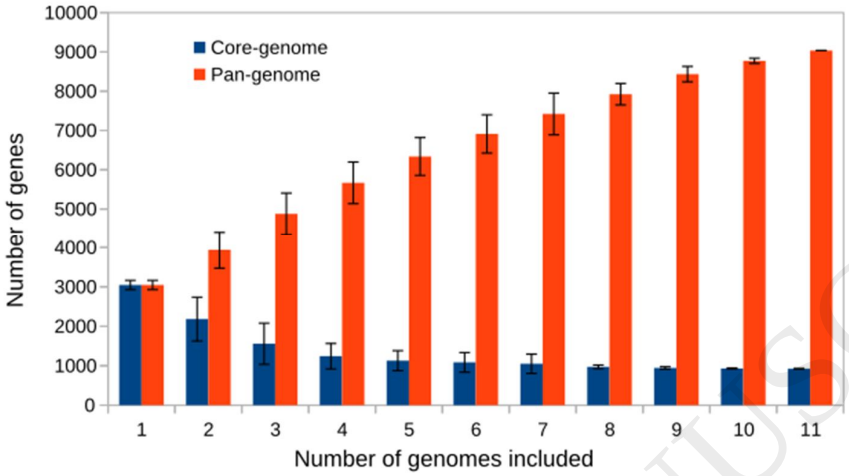


Figure 4: Venn diagrams showing the composition of unique and shared genes at different genomic levels: (A) all four *Sal. ruber* genomes; (B) Argentinian isolates, (C) the two *Slv.* species, (D) EHB-1 of *Sal. ruber* and the Argentinian strains, (E) *Sal. ruber* EHB-1 and the *Slv.* species, and (F) *Sal. ruber* EHB-1, *Sat. longa* and *Rho. marinus* genomes.

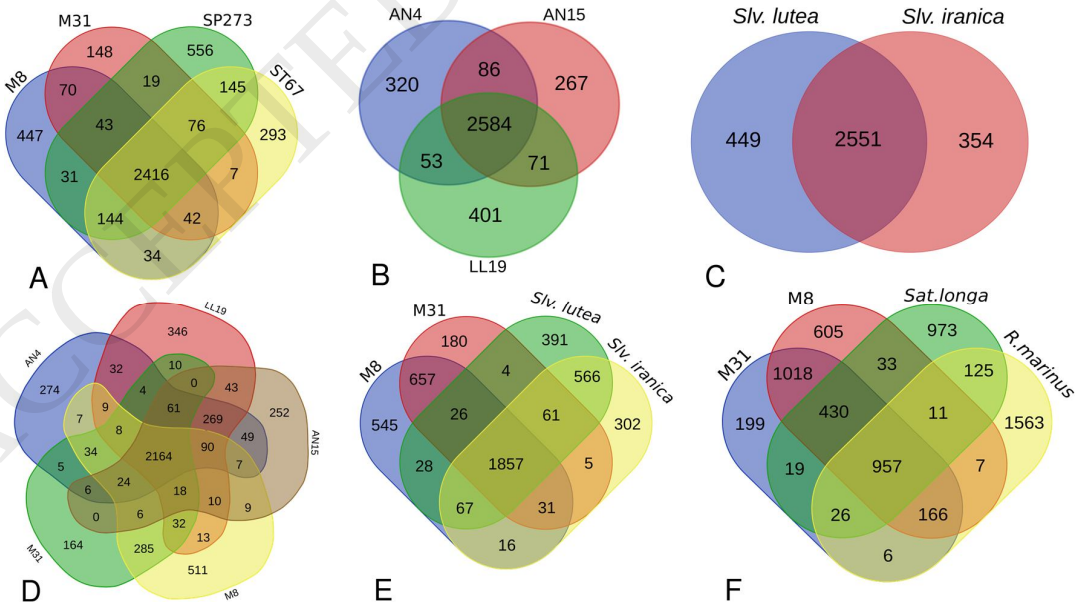


Figure 5: Hierarchical clustering based on the presence (yellow) or absence (blue) of a total of 3,306 variable genes (genes shared between two or more genomes). Pan-genome of the members of *Salinibacteraceae* (A), *Salinivenus* and *Salinibacter* (B), the *Salinibacter* genus (C), *Sal. ruber* (D), *Salinivenus* species (E), *Sal. ruber* EHB-2 (F), the Argentinian strains (F), and *Sal. ruber* EHB-2 (H).

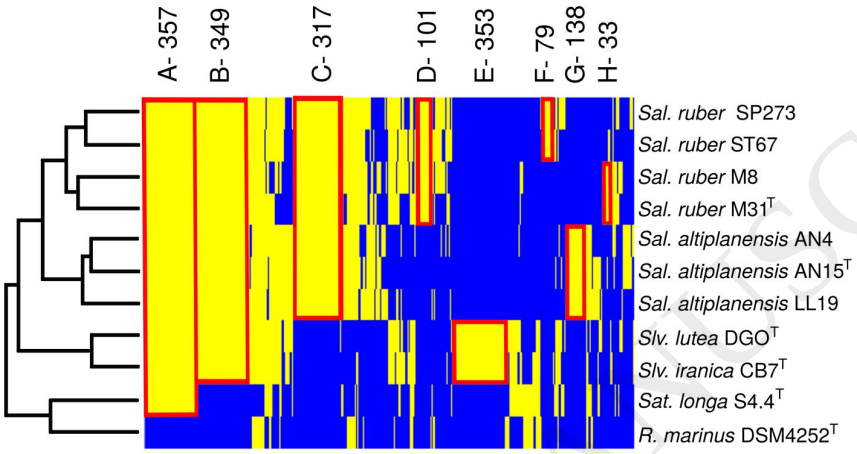


Figure 6: Synteny between genomes for the family *Salinibacteraceae* using M31^{T} as the reference.

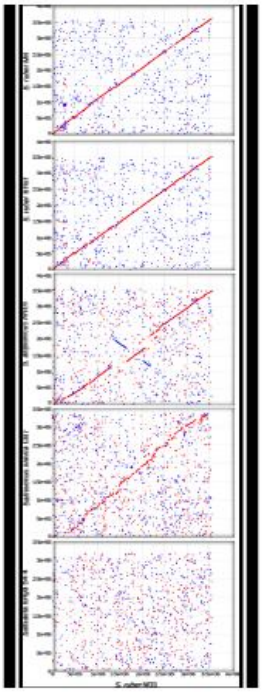
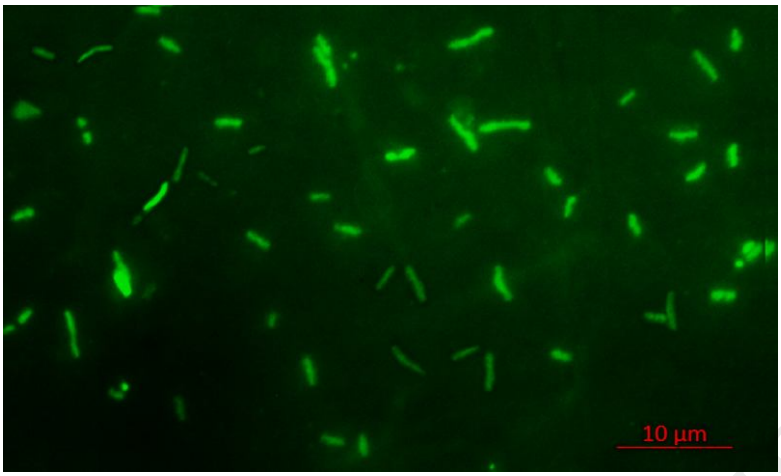


Figure 7: CARD-FISH of *Sal. altiplanensis*.



Tables

Table 1: Strains in the study, origin and most relevant genomic and phenotypic features.

	Sal. ruber				*Sal. altiplanensis*			*Slv. iranica*	*Slv. lutea*
Strain	**M8**	**M31^T**	**ST67**	**SP273**	**AN15^T**	**AN4**	**LL19**	**CB7^T**	**DGO^T**
Country of isolation	Spain	Spain	Spain	Spain	Argentina	Argentina	Argentina	Iran	Iran
Region	Mallorca	Mallorca	Alicante	Tarragona	Antofalla	Antofalla	Llullialliaco	Aran-Bidgold	Aran-Bidgold
Latitude	39º21'05'' N	39º21'05'' N	38º12'06'' N	40º35'60''N	25°37'18"S	25°37'18"S	24°48'0"S	34º30'8.11''N	34º30'8.11''N
Longitude	3º00'21''E	3º00'21''E	0º35'47'' O	0º41'24''E	67°38'30"W	67°38'30"W	68°17'24"W	51°45'50.04" E	51°45'50.04" E
Altitude (asl, m)	0	0	0	0	4000	4000	3677	800	800
Year of isolation	2000	2000	2010	2010	2012	2012	2012	2007	2007
Sampling date	Sep, 1999	Sep, 1999	Jan, 2011	Jan, 2011	Feb, 2011	Feb, 2011	Feb, 2011	Nov, 2007	Nov, 2007
G+C mol% (from genome data)	65.85	66.14	66.22	65.49	64.41	64.24	44.38	65.66	65.9
G+C mol% (HPLC)	66.5	67.0	nd	nd	nd	nd	nd	64.8	65.6
Cell size (µm)	4.6 ± 0.7	5.5 ± 1.5	5.1 ± 1.1	5.9 ± 1.4	4 ± 0.84	3.2 ± 0.9	3.7 ± 1.17	10 ± 2.5	7.1 ± 2.2
Colony pigmentation	Red	Red	Red	Red	Red	Red	Red	Red	Orange
Motility	+	+	+	+	+	+	+	-	-
Optimum salt (%)	20	20	20	20	25	25	25	20	20
Optimum pH	7	7	7	7	7.5	7.5	7.5	7.5	7.5
Optimum T (ºC)	40	40	40	40	30	30	30	35	35
Gelatin hydrolysis	+	+	+	+	-	-	-	-	-
Starch hydrolysis	+	+	+	+	+	+	+	-	-
16S rRNA gene acc. no.	AF323501	AF323500	LN650034	LN650035	LT160741	LT160743	LT160742	HQ197982	HQ197983
Genome acc. no.	FP565814.1	CP000159.1	CP020719	NCQX00000000	NCQY00000000	NCRA00000000	NCQZ00000000	NCRB00000000	NCRC00000000
Strain collection no.		DSM 13855 = CECT 5946	CECT 9123	CECT 9122	IBRC-M 11031 = CECT 9105	IBRC-M 11030 = CECT 9106	IBRC-M 11032	CGMCC 1.1100 = IBRC-M 10036	CGMCC 1.11002 = IBRC-M 10423

Table 2: Genomes in the study.

Strain	Phylotype	Sampling site	Status	Sequencing method	Assembler	Coverage	Num. Ctg.	Genome size (Mb)	%GC	Predicted CDS	rRNA proteins	tRNAs	Hypothetical proteins	Genome acc. no. (GenBank sequence)
M31^T	PH1[a]	S'Avall Salterns (Mallorca; Spain)	Complete		TIGR		2	3.59	66.14	3,152	53	55	1,158	CP000159.1
M8	PH1[b]	S'Avall Salterns (Mallorca; Spain)	Complete	ABI3730	Phred/Phrap/Consed		5	3.83	65.85	3,321	52	56	1,269	FP565814.1
ST67	PH2	Trinitat Salterns (Tarragona; Spain)	Complete	PacBio and Roche GS FLX	Spades and SSPACE	36x	3	3.5	66.22	3,258	52	56	1,188	CP020719*
SP273	PH2	Santa Pola Salterns (Alicante; Spain)	Draft	Illumina Miseq	IDBA	77x	219	4.06	65.49	3,362	52	55	1,301	NCQX00000000*
AN15^T	PH3	Salar de Antofalla (Argentina)	Draft	Illumina Hiseq	Velvet	219x	30	3.58	64.41	3,135	53	58	1,154	NCQY00000000*
AN4	PH3	Salar de Antofalla (Argentina)	Draft	Illumina Miseq	IDBA	115x	95	3.65	64.24	3,171	52	59	1,147	NCRA00000000*
LL19	PH3	Salar de Llullaillaco (Argentina)	Draft	Illumina Hiseq	Velvet	366x	52	3.71	64.38	3,222	53	60	1,207	NCQZ00000000*
CB7^T	*S. iranica*[c]	Aran-Bigdol Lake (Iran)	Draft	Illumina Miseq	IDBA	129x	133	3.41	65.66	2,931	53	58	947	NCRB00000000*
DGO^T	*S. lutea*[c]	Aran-Bigdol Lake (Iran)	Draft	Illumina Miseq	IDBA	255x	76	3.55	65.9	3,024	53	59	1,011	NCRC00000000*
S4-4^T	*S. longa*[d]	Mixture of water from Dead Sea and Red Sea (Israel)	Complete	Illumina	Velvet	Unknown	3	3.4	63.5	2,635	57	47	1,016	GCA_000419585.1
DSM4252	*R. marinus*[e]	NW Iceland	Complete	Sanger and 454 GS FLX	Newbler	23.8x	2	3.26	64.5	2,815	57	45	772	CP001807.1

a) Mongodin et al., 2005 [39]; b) Peña et al., 2010 [51]; c) Makhdoumi-Kakhki et al., 2012 [36]; d) Vaisman et al., 2009 [68]; e) Nolan et al., 2009 [44]

* Genome sequences generated in this study

Table 3: CRISPR spacers found in the different genomes sequenced. Note that the pair of spacers LL19_3 and AN4_2 (dark grey shaded), and LL19_4 and AN4_2 (light grey shaded) were identical.

Species	Strain	CRISPR spacer information					
		Number	Contig	Number repetitions	Start position	End position	Spacer sequence
Sal. ruber	ST67	CRISPR – 1	1	4	1884673	1884945	GGGTCTATCTCCGTGTGTGCGGAGGTACC
	SP273	CRISPR – 1	65	12	47581	48419	CCTCTAATCGCACCTTTGAGGTATTGAAAG
		CRISPR – 2	65	37	52252	54735	CTCCTAATCGCACCTTTGAGGTATTGAAAG
Sal. altiplanensis	AN15^{T}	CRISPR – 1	88	9	575628	576205	GGTAACTCCGCACACGCGGAGATAGACC
		CRISPR – 2	88	14	587940	588822	GGTAGCCCCGCACTCGCGGGGATAGACCC
	LL19	CRISPR – 1	5	3	584	795	GGTAGCTCCGCGCATGCGGGGATAGTCC
		CRISPR – 2	55	5	123183	123501	GGTGGCCCCGCATGCGCGGGGATAGTCC
		CRISPR – 3	55	12	159135	159894	GGTCTATCCCCGCGAGTGCGGGGCTACC
		CRISPR – 4	55	14	171631	172512	GGTCTATCTCCGCGTGTGCGGAGTTACC
	AN4	CRISPR – 1	2	48	108888	111843	GGTCTATCCCCGCGAGTGCGGGGCTACC
		CRISPR – 2	2	24	129881	131373	GGTCTATCTCCGCGTGTGCGGAGTTACC
		CRISPR – 3	12	3	60734	60930	GGGAGCGAGCTGCGCCCAAAGTCCGG
Slv. lutea		CRISPR – 1	0	10	67271	67908	GGTAGCCCCGCATCCGCGGGGATAGACC
		CRISPR – 2	2	9	64894	65470	GGTGCCCCCGCAATCCGCGGGGATAGACC
		CRISPR – 3	5	9	33958	34591	CTTCTAATCGCACCTTTGAGGTATTGAAAG
		CRISPR – 4	6	6	149337	149726	GGTGACTCCGCACACGCGGAGATA
		CRISPR – 5	26	9	42977	43553	GGTCTATCTCCGCGTGTGCGGAGGCATC
Slv. iranica		CRISPR – 1	2	19	136034	137329	GTAGCATCCCCCCTCACCGGGGGGATGAGGATTGAAAG
		CRISPR – 2	7	3	41267	41477	GGTAGCCCCGCATCCGCGGGGATAGACCC
		CRISPR – 3	12	9	52053	52629	GGTCTATCCCCGCGGATGCGGGGGCACC
Salisaeta longa		CRISPR – 1	Chr.	5	56721	57143	GTCGGAAGTCTTGCCTCCATTCCAAGAGGATTGAAAC
		CRISPR – 2	Chr.	5	410382	410743	ATTTCAATACCTCAAAGGTGCGATTAAAAG
		CRISPR – 3	Chr.	6	918760	919272	GTTTCAATCCTCTTGGAATGGAGGCAAGACTTCCGAC
		CRISPR – 4	Chr.	4	940600	940952	GTTTCAATCCTCTTGGAATGGAGGCAAGACTTCCGAC
		CRISPR – 5	Chr.	8	993655	994214	ATTTCAATACCTCAAAGGTGCGATTAAAAG
		CRISPR – 6	Chr.	30	1577747	1579611	GGTCTATCCCCGCGTGTGCGGGGTCATC
		CRISPR – 7	Chr.	20	3022280	3023527	GGTCTATCCCCGCGTGTGCGGGGTCATC
		CRISPR – 8	Chr.	4	1919263	1919608	GTCGGAAGTCTTGCCTCCATTCCAAGAGGATTGAAAC
		CRISPR – 9	Chr.	7	3178174	3178761	GTCGGAAGTCTTGCCTCCATTCCAAGAGGATTGAAAC
		CRISPR – 10	Chr.	4	3182381	3182732	GTCGGAAGTCTTGCCTCCATTCCAAGAGGATTGAAAC
Rhodothermus marinus		CRISPR – 1	Chr.	73	539686	544167	GGTGTCCCCGCACCCGCGGGGATAGTCCC
		CRISPR – 2	Chr.	43	1886877	1889769	ATTTCAATACCAAAAAGGTGCGATTAAAAC

	CRISPR – 3	Chr.	38	2079047	2081608	ATTTCAATACCAAAAAGGTGCGATTAAAAC
	CRISPR – 4	Chr.	8	24096	24759	GTCGTAATCCCCTTTTCATCGGGTCAAGTCTTCGGAC
	CRISPR – 5	Chr.	11	32202	33116	GTCGTAATCCCCTTTTCATCGGGTCAAGTCTTCGGAC
	CRISPR – 6	Chr.	6	55091	55603	GTCGTAATCCCCTTTTCATCGGGTCAAGTCTTCGGAC
	CRISPR – 7	Chr.	6	65829	66347	GTCGTAATCCCCTTTTCATCGGGTCAAGTCTTCGGAC
	CRISPR – 8	Chr.	18	117160	118505	CCTTCAATGCTGCCGTAGCTATTTAGCTACGGAAAT
	CRISPR – 9	Chr.	32	119724	122080	CCTTCAATGCTGCCGTAGCTATTTAGCTACGGAAAT

Table 4: Description of *Salinibacter altiplanensis* sp. nov. according to the digitalized protologue TA00140 at the www.imedea.uib.es/dprotologue website

TAXONUMBER	TA00140
SPECIES NAME	*Salinibacter altiplanensis*
GENUS NAME	*Salinibacter*
SPECIFIC EPITHET	*altiplanensis*
SPECIES STATUS	sp. nov.
SPECIES ETYMOLOGY	al.ti.pla.ne'nsis, N.L. masc. adj. *altiplanensis* of the Argentinian altiplano
AUTHORS	Viver T, Orellana L, Gonzalez P, Diaz S, Urdiain M, Farias ME, Benes V, Kaempfer P, Shahinpei A, Amoozegar M, Amann R, Anton J, Konstantinidis KT, Rossello-Mora R
TITLE	Genomic comparisons between members of the *Salinibacteraceae* family, and classification of a new species of *Salinibacter* (*S. altiplanensis* sp. nov.) isolated from high altitude hypersaline environments of the Argentinian Altiplano
CORRESPONDING AUTHOR	Ramon Rosselló-Móra
E-MAIL OF THE CORRESPONDING AUTHOR	ramon@imedea.uib-csic.es
SUBMITTER	RAMON ROSSELLÓ-MÓRA
E-MAIL OF THE SUBMITTER	ramon@imedea.uib-csic.es
DESIGNATION OF THE TYPE STRAIN	AN15
STRAIN COLLECTION NUMBERS	IBRC-M 11031 = CECT 9105
16S rRNA GENE ACCESSION NUMBER	LT160741
GENOME ACCESSION NUMBER [RefSeq]	NCQY00000000
GENOME STATUS	Complete
GENOME SIZE	3580
GC mol%	64.41
COUNTRY OF ORIGIN	Argentina
REGION OF ORIGIN	Salar de Antofalla
DATE OF ISOLATION	5/7/12
SOURCE OF ISOLATION	Hypersaline lake
SAMPLING DATE	20/2/11
GEOGRAPHIC LOCATION	Salar de Antofalla
LATITUDE	25º37'18"S
LONGITUDE	67º38'30"W
ALTITUDE	4000
NUMBER OF STRAINS IN STUDY	3
SOURCE OF ISOLATION OF NON-TYPE STRAINS	Hypersaline lakes
GROWTH MEDIUM, INCUBATION CONDITIONS [Temperature, pH, and further information] USED FOR STANDARD CULTIVATION	Sea water 25% salt concentration [Rodríguez-Valera et al. (1985) 11: 107-115]

IS A DEFINED MEDIUM AVAILABLE	Sea water 25% salt concentration [Rodríguez-Valera et al. (1985) 11: 107-115]
ALTERNATIVE MEDIUM 1	Medium A [Antón J et al. (2002) 52: 485-491]
ALTERNATIVE MEDIUM 2	Medium B [Antón J et al. (2002) 52: 485-491]
CONDITIONS OF PRESERVATION	Liquid medium (SW 20%) mixed with 40% (v/v) glycerol and stored at -80 ºC
GRAM STAIN	NEGATIVE
CELL SHAPE	Long rod
CELL SIZE (length or diameter)	4 ± 0.84
MOTILITY	Motile
IF MOTILE	Flagellar
COLONY MORPHOLOGY	Colonies were red, approximately 1 mm in diameter, circular and convex with an entire margin
TEMPERATURE OPTIMUM	30
pH OPTIMUM	7.5
pH CATEGORY	Neutrophile
SALINITY OPTIMUM	25
SALINITY CATEGORY	Extreme halophilic (optimum >15% NaCl)
RELATIONSHIP TO O_2	Aerobe
O_2 CONDITIONS FOR STRAIN TESTING	Aerobiosis
CARBON SOURCE USED [class of compounds]	Sugars, amino acids
CARBON SOURCE USED [specific compounds]	D-fructose, D-glucose, sucrose
CARBON SOURCE NOT USED [specific compounds]	D-ribose, lactose, L-alanine, L-asparagine, L-aspartate, L-cysteine, L-methionine, L-phenylalanine, L-tyrosine
CARBON SOURCE VARIABLE [specific compounds]	D-galactose, L-lysine, D-mannitol, L-proline, L-tryptophan, maltose, raffinose
ENERGY METABOLISM	Chemoorganotroph
OXIDASE	Positive
CATALASE	Positive
POSITIVE TESTS	DNA hydrolysis, lysine decarboxylase, ornithine decarboxylase
NEGATIVE TESTS	Anaerobic growth with arginine or DMSO, hydrolysis of Tween 20, casein hydrolysis, arginine hydrolase, indole production, Voges-Proskauer test, methyl red test, production of H2S, gas formation with nitrate
MAJOR FATTY ACIDS	$C_{15:0}$ iso; $C_{15:0}$ anteiso; $C_{16:0}$; $C_{17:0}$ iso 3-OH; $C_{18:1}$ ω7c
BIOSAFETY LEVEL	1
HABITAT	Saline evaporation pond (ENVO:00000055)
BIOTIC RELATIONSHIP	Free-living
KNOWN PATHOGENICITY	None