The genomic sequence of *Exiguobacterium chiriqhucha* str. N139 reveals a species that thrives in cold waters and extreme environmental conditions

Ana Gutiérrez-Preciado [Corresp., 1, 2], **Carlos Vargas-Chávez** [1], **Mariana Reyes-Prieto** [1], **Omar F Ordoñez** [3], **Diego Santos-García** [1, 4], **Tania Rosas-Pérez** [1], **Jorge Valdivia-Anistro** [5, 6], **Eria A Rebollar** [7], **Andrés Saralegui** [8], **Andrés Moya** [1], **Enrique Merino** [9], **María Eugenia Farías** [10], **Amparo Latorre** [Corresp., 1], **Valeria Souza** [Corresp. 6]

[1] Unidad de Genética Evolutiva, Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universidad de Valencia, Calle Catedrático José Beltrán Martínez, Paterna, Valencia, Spain

[2] Ecologie Systématique Evolution, CNRS, AgroParisTech, Université Paris Sud (Paris XI), Orsay, France

[3] Laboratorio de Investigaciones Microbiológicas de Lagunas Andinas, Planta Piloto de Procesos Industriales Microbiológicos (PROIMI), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Av. Belgrano y Pasaje Caseros, San Miguel de Tucumán, Argentina

[4] Department of Entomology, Hebrew University of Jerusalem, Rehovot, Israel

[5] Carrera de Biología, Faculta de Estudios Superiores Zaragoza, UNAM, Mexico City, Mexico

[6] Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, coyoacan, Mexico City, México

[7] Department of Biology, James Madison University, Harrisonburg, Virginia, United States of America

[8] Laboratorio Nacional de Microscopía Avanzada, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México

[9] Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México

[10] Laboratorio de Investigaciones Microbiológicas de Lagunas Andinas, Planta Piloto de Procesos Industriales Microbiológicos (PROIMI), Consejo Nacional de Investigaciones Científicas y Técnicas, Av. Belgrano y Pasaje Caseros, San Miguel de Tucumán, Argentina

Corresponding Authors: Ana Gutiérrez-Preciado, Amparo Latorre, Valeria Souza
Email address: anagtz@gmail.com, Amparo.Latorre@uv.es, souza.valeria2@gmail.com

We report the genome sequence of *Exiguobacterium chiriqhucha* str. N139, isolated from a high-altitude Andean lake. Comparative genomic analyses of the *Exiguobacterium* genomes available suggest that our strain belongs to the same species as the previously reported *E. pavilionensis* str. RW-2 and *Exiguobacterium* str. GIC 31. We describe this species and propose the *chiriqhucha* name to group them. ‘Chiri qhucha’ in quechua means ‘cold lake’, which is a common origin of these three cosmopolitan Exiguobacteria. The 2,952,588-bp *E. chiriqhucha* str. N139 genome contains one chromosome and three megaplasmids. The genome analysis of the Andean strain suggests the presence of enzymes that confer *E. chiriqhucha* str. N139 the ability to grow under multiple environmental extreme conditions, including high concentrations of different metals, high ultraviolet B radiation, scavenging for phosphorous and coping with high salinity. Moreover, the regulation of its tryptophan biosynthesis suggests that novel pathways remain to be discovered, and that these pathways might be fundamental in the amino acid metabolism of the microbial community from Laguna Negra, Argentina.

The genomic sequence of *Exiguobacterium chiriqhucha* str. N139 reveals a species that thrives in cold waters and extreme environmental conditions

Ana Gutiérrez-Preciado[1†‡], Carlos Vargas-Chávez[1†], Mariana Reyes-Prieto[1], Omar F Ordoñez[2], Diego Santos-Garcia[1ρ], Tania Rosas-Pérez[1], Jorge Valdivia-Anistro[3±], Eria A Rebollar[4], Andrés Saralegui[5], Andrés Moya[1], Enrique Merino[6], María Eugenia Farias[2], Amparo Latorre[1*] and Valeria Souza[3*].

[*]Corresponding authors: Amparo Latorre E-mail: amparo.latorre@uv.es; Valeria Souza E-mail: souza@servidor.unam.mx

[†]Equal contributors

[1]Unidad de Genética Evolutiva, Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universidad de Valencia, Calle Catedrático José Beltrán Martínez 2, 46980, Paterna, Spain

[2]Laboratorio de Investigaciones Microbiológicas de Lagunas Andinas (LIMLA), Planta Piloto de Procesos Industriales Microbiológicos (PROIMI), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Av. Belgrano y Pasaje Caseros, 4000 San Miguel de Tucumán, Argentina

[3]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, México D.F., México

[4]Department of Biology, James Madison University, Harrisonburg, Virginia, 22801, United States of America

[5]Laboratorio Nacional de Microscopía Avanzada, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México

[6]Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apdo. Postal 510-3, Cuernavaca, Morelos, México

[‡]Present address: Ecologie Systématique Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, Orsay, France

[ρ]Present addresss: Department of Entomology, Hebrew University of Jerusalem, Rehovot, Israel

[±]Present address: Carrera de Biología, Faculta de Estudios Superiores Zaragoza, UNAM, Mexico City, Mexico

Data deposition: This Whole Genome Shotgun project has been deposited at GenBank under the accession JMEH00000000. The version described in this paper is version JMEH001000000.1; BioSample SAMN02732272

Abstract

We report the genome sequence of *Exiguobacterium chiriqhucha* str. N139, isolated from a high-altitude Andean lake. Comparative genomic analyses of the *Exiguobacterium* genomes available suggest that our strain belongs to the same species as the previously reported *E. pavilionensis* str. RW-2 and *Exiguobacterium* str. GIC 31. We describe this species and propose the *chiriqhucha* name to group them. 'Chiri qhucha' in quechua means 'cold lake', which is a common origin of these three cosmopolitan Exiguobacteria. The 2,952,588-bp *E. chiriqhucha* str. N139 genome contains one chromosome and three megaplasmids. The genome analysis of the Andean strain suggests the presence of enzymes that confer *E. chiriqhucha* str. N139 the ability to grow under multiple environmental extreme conditions, including high concentrations of different metals, high ultraviolet B radiation, scavenging for phosphorous and coping with high salinity. Moreover, the regulation of its tryptophan biosynthesis suggests that novel pathways remain to be discovered, and that these pathways might be fundamental in the amino acid metabolism of the microbial community from Laguna Negra, Argentina.

Short title: Genome of *E. chiriqhucha* str. N139

Abbreviations

HAALs: high altitude Andean Lakes ANI: Average Nucleotide Identity AAI: Average Amino Acid Identity

LM: Lake Medium

SSGs: Strain Specific Genes HGT: Horizontal Gene Transfer

BCAA: Branched Chain Amino Acids PER: Photoenzymatic Repair

NER: Nucleotide Excision Repair PRR: Post Replication Repair

The high altitude Andean Lakes (HAALs) from Puna, Argentina, are a group of lakes located at 3000-6000 meters above sea level which are characterized by high ultraviolet (UV) radiation and salinity, broad temperature variations, low nutrient concentrations and high contents of metals and metalloids, mainly arsenic (Fernández-Zenoff et al., 2006; Fernández-Zenoff, Siñeriz & Farías, 2006; Dib et al., 2008; Flores et al., 2009; Ordoñez et al., 2009; Albarracín et al., 2011; Belfiore, Ordoñez & Farías, 2013). These environmental conditions are considered to be extreme and might resemble those of the Earth's early atmosphere, as has been stated by NASA (Cabrol et al., 2007; Farías et al., 2009). Hence, these geographical areas have been proposed for studies on astrobiology (Farías et al., 2009). Despite being oligotrophic and hostile, a great microbial diversity has been found in the HAALs, where bacteria from the genus Exiguobacterium are one of the dominant taxa (Ordoñez et al., 2009, 2013; Sacheti et al., 2013).

The Exiguobacterium genus, a sister clade to the Bacillus genus, is currently underexplored, and molecular studies of this genus from different sources are limited (Vishnivetskaya, Kathariou & Tiedje, 2009). Exploring *Exiguobacterium* strains is of great significance because understanding their strategies to adapt to diverse and extreme environmental conditions will likely place them as model organisms involved in the remediation of organic and inorganic pollutants. In particular, *Exiguobacterium* strains isolated from the HAALs have the potential of becoming an attractive model system to study environmental stress responses, as these microorganisms are able to grow efficiently in the laboratory (Ordoñez et al., 2009; Belfiore, Ordoñez & Farías, 2013). Moreover, Dib *et al.* suggested that these microorganisms could harbor various stress defense associated systems (Dib et al., 2008).

The *Exiguobacterium chiriqhucha* str. N139 was selected for genome sequencing due to its stress defense mechanisms such as its tolerance to high UV-B radiation, salinity and metalloids, particularly arsenic. This strain was isolated from the water column of Laguna Negra, which belongs to the 'Salar de la Laguna Verde', a system of five shallow oligotrophic lakes originated in the Tertiary (65 million to 1.8 million years ago) (Ericksen & Salas, 1987).

In the present study we characterized the genome of *E. chiriqhucha* str. N139, in order to identify the strategies that this organism employs to cope with the extreme environmental factors present in the aforementioned lake, mainly those related to metal and UV-B resistance. We also performed comparative genomics focusing on the two strains that would comprise the same species as *Exiguobacterium chirqhucha* str. N139; *E. pavilionensis* str. RW-2, isolated from the permanently cold Pavilion Lake in Canada (White III, Grassa & Suttle, 2013) and *Exiguobacterium* str. GIC31 isolated from a glacier in Greenland (Vishnivetskaya et al., 2014).

Since the three strains were isolated from cold lakes, we propose the name 'chiri qhucha', which means 'cold lake' in quechua, the Andean prehispanic language.

Classification and features

Members of the Exiguobacterium genus are Firmicutes, Gram-positive, facultative anaerobes with a low G + C content (Vishnivetskaya, Kathariou & Tiedje, 2009). *Exiguobacterium* is widely distributed all over the world (Karami et al., 2011) and has been isolated and typified from a wide variety of environments including hot springs (Vishnivetskaya, Kathariou & Tiedje, 2009; Vishnivetskaya et al., 2011), hydrothermal vents (Crapart et al., 2007), permafrost (Vishnivetskaya & Kathariou, 2005; Vishnivetskaya et al., 2006; Rodrigues et al., 2008), marine sediment (Kim et al., 2005), oligotrophic environments (Rebollar et al., 2012), biofilms (Carneiro et al., 2012), alkaline methanogenic microcosms (Rout, Rai & Humphreys, 2015) and more recently in water and microbial mats from high-altitude desert wetlands (Ordoñez et al., 2013). The Exiguobacterium genus is divided in two main phylogenetic clades (Vishnivetskaya, Kathariou & Tiedje, 2009); clade I is composed of temperate and cold-adapted strains, whereas clade II includes alkaliphilic species, with a marine origin and/or from high-temperature habitats (Figure 1A).

E. chiriqhucha str. N139, which belongs to clade II, was isolated from the water column of Laguna Negra, in the HAALs (GPS: 27°38'49" S, 68°32'43" W) and in laboratory conditions can uptake a wide variety of carbon sources (Table S1). Its cells are short rods and do not sporulate (Figure 2, Table 1).

115 ## Materials and Methods

116

117

118 ### Growth conditions and genomic DNA preparation

119 *E. chiriqhucha* str. N139 was isolated from Laguna Negra by plating it in Lake Medium
120 (LM). LM was used to maintain the same salinity as the isolation environment and was obtained
121 by filtering lake water (0.22 µm Biopore filters) and adding 2.5 g of yeast extract and 12 g of agar
122 (Difco) per liter at 20 °C. For future assays the strain was grown in LM broth at 20 °C with
123 agitation. DNA was extracted using the protocol described by (Fernández-Zenoff, Siñeriz &
124 Farías, 2006).

125 ### Microscopy

126 Differential interference contrast (DIC) images were obtained from cells grown on LB
127 medium overnight, and mounted in No. 2 coverslips (Figure 2). LB medium was used as mounting
128 media during image acquisition. Images were shot with an Olympus FV1000 Laser Scanning
129 Confocal on an Olympus IX81 inverted microscope equipped with 60x UPlanSApo NA 1.3 Sil
130 objective lens. With a 405nm laser line, DIC Images were acquired in the TD channel controlled
131 with Olympus FV10-ASW-4.2 software. Brightness, contrast and scale bars were adjusted on
132 displayed images using the Fiji software.

133 ### Phylogenetic Reconstruction

134 The complete genomic sequences of 17 representative *Exiguobacterium* strains were used
135 to reconstruct their phylogeny with PhyloPhlAn. This software extracts a set 31 manually curated

conserved proteins from each genome, aligns them, keeps the positions that retain important evolutionary information and builds a phylogenetic tree (Segata et al., 2013). Figure 1A shows this phylogenetic reconstruction of the selected *Exiguobacterium* genomes from both clades.

Genome sequencing and assembly

The genome of *E. chiriqhucha* str. N139 was generated using *454 technology* (Table 2). A standard *454 Titanium* library was constructed and sequenced, producing 664,086 reads, totaling 253.9 Mb of data. Phred quality cut-off was set to 20. The 454 data was assembled with *Newbler*, version 2.8 and MIRA, version 3.4 (Chevreux et al., 2004). The GS De Novo Assembler GUI was used, parameters selected were minimum read length of 45 and output scaffolds file. The parameters used for the MIRA assembly were: number of passes of 1 and no uniform read distribution nor trimming of overhanging reads. Warning message for read names longer than 40 was deactivated. Both assemblies were merged using *Minimus2*, from the *amos* version 3.1.0, with assembly errors manually corrected. The contigs were sorted with *Mauve* version 2.3.1 (Rissman et al., 2009), using *Exiguobacterium* sp. AT1b as the reference because it is the closest relative with a completely sequenced genome (Vishnivetskaya et al., 2011).

Genome annotation

Protein-coding genes, tRNAs, rRNAs and non-coding RNAs were identified using the annotation pipeline *Prokka* (Seemann, 2014), followed by annotation refinement with *InterProScan* (Quevillon et al., 2005). Riboswitches were identified with the *Infernal* 1.1 package (Nawrocki & Eddy, 2013) using the corresponding covariance models from the *Rfam* database (Burge et al., 2013). COGs were assigned by profile hidden Markov model (profile HMM) searches using the *hmmsearch* program of the *HMMER3* package (Mistry et al., 2013). For every

158 COG, a multiple sequence alignment of *bona fide* representative sequences were generated using
159 the *Muscle* program (Edgar, 2004), and then, the corresponding Hidden Markov Model was built
160 using the *hmmerbuild* program, also provided in the *HMMER3* package (Mistry et al., 2013). The
161 cutoff E-value in the *hmmsearch* process varies importantly for every COG. For every one of the
162 COG groups we have defined a high confidence cutoff E-value value defined as the highest E-
163 value (smallest bit score) observed for the members of such COG. In any case, none of the COG
164 cutoff E- values was greater than 1e-10.The resulting annotation was subjected to manual curation.
165 *Pathway Tools 13 (Karp et al., 2009)* in combination with the *BioCyc* (Caspi et al., 2014) and
166 *UNIPROT* (Consortium, 2015) databases were used to infer the metabolic capacities of *E.*
167 *chiriqhucha* str. N139. The curated model of *E. chiriqhucha* str. N139 can be provided upon
168 request, and will be deposited in the *BioCyc* database.

169 ## Clustering of strains into a single species

170 Phylogenetic analyses, ANI and AAI calculations, synteny analyses, as well as pangenome
171 reconstruction were performed in order to a better understanding of the taxonomic position of the
172 strain N139 isolated from Laguna Negra, relative to all the *Exiguobacterium* genomes available at
173 the time of analysis.

174 ANI and AAI calculations (Goris et al., 2007) were done with default parameters for N139
175 *versus* all other complete genomic sequences of *Exiguobacterium* as well as pairwise comparisons
176 for the three closest strains using the web server from Kostas lab http://enve-omics.ce.gatech.edu/
177 with default parameters. Comparisons to *Exiguobacterium* str. N139 can be seen in Table S2.

178 To explore the genomic rearrangements present on *E. chiriqhucha* str. N139 in comparison
179 to other *Exiguobacterium* species, nucleotide syntenic blocks were obtained with *Mauve* version

2.3.1 (Darling et al., 2004). Syntenic block permutations were exported and used as input for *MGR* (Bourque & Pevzner, 2002). *MGR* was used to calculate the minimum number of rearrangements between the species analyzed, and to recover the rearrangement dendrogram. *genoPlotR* (Guy, Kultima & Andersson, 2010) was used to plot the syntenic blocks (Figure 1B).

The pangenome of the nine *Exiguobacterium* genomes from clade II available at the time was reconstructed to aid in the taxonomic positioning of the strain N139. Orthologs were first calculated following the OrthoMCL pipeline (Li, Jr & Roos, 2003; Fischer et al., 2011), and the pangenome and the core genome were elucidated using *ad hoc* perl scripts. The selected strains were *E*. str. AT1b isolated from a hot spring in Yellowstone, USA (Vishnivetskaya et al., 2011); *E. marinum*, isolated from the Yellow Sea, South Korea (Vishnivetskaya et al., 2014); *E. aurantiacum*, from a potato processing plant, in the UK (Vishnivetskaya et al., 2014); *E.* str. 8-11-1 isolated from a salt lake in Inner Mongolia, China (Jiang et al., 2013); *E*. sp. S17 from the Laguna Socompa, another HAAL, Argentina (Ordoñez et al., 2013); *E. mexicanum* isolated from a brine shrimp *Artemia franciscana (López-Cortés et al., 2006)*; *E. pavilionensis* (now *chiriqhucha*) str. RW-2 isolated from Pavilion Lake, Canada (White III, Grassa & Suttle, 2013) and *E.* (now *chiriqhucha*) str. GIC31, isolated from glacier ice in Greenland (Vishnivetskaya et al., 2014).

UV resistance assays: determination of survival rate

The strains *Exiguobacterium* sp. S17, isolated from Laguna Socompa (HAAL), and *Exiguobacterium aurantiacum* str. DSM 6208, from the German Collection of Microorganisms and Cell Cultures (DSM), were used in these assays for comparison as external controls. These

strains and *Exiguobacterium* str. N139 were grown in 40 mL of LB medium under shaking (150 rpm) at 30°C and cells were harvested by mid log phase (OD_{600nm} 0.5) by centrifugation at 8000 rpm for 10 min at 4ºC. The pellets were washed twice with 30 mL of 0.9% NaCl, and resuspended in the same volume of 40 mL. 20 mL of cell suspensions were transferred into sterile quartz tubes (16 cm long and 1.8 cm diameter) and placed horizontally to ensure maximal exposure and incubated at 15°C under gentle shaking (150 rpm).

Tubes were irradiated from a distance of 30 cm with UV-B doses between 2,0 - 3,0 W/m2 during 240 min (09815-06 lamps, Cole Parmer Instrument Company; major emission line at 312 nm). Tubes were covered with an acetate sheet to block out UV-C. Irradiance was quantified with a radiometer (09811-56, Cole Parmer Instrument Company) at 312 nm with half bandwidth of 300 to 325 nm. Aliquots of 0.1 mL were taken at different exposure times (0, 60, 120, 180 and 240 min). Samples were then serially diluted in LB broth and spread in duplicate on Petri dishes with the same medium to determine the number of colony forming units (CFU). Controls of unexposed samples were run simultaneously in darkness and the percentage of cell survival after each treatment was calculated relative to these controls.

Results and Discussion

Genome Properties

The final assembly of the genome of *E. chiriqhucha* str. N139 consists of 23 contigs, the smallest one being 457 bases in length and the largest 1.5 Mb, with an average coverage of 85×. Its genome includes three circular megaplasmids with probable sizes of 250.57, 137.48 and 48 Kb, as determined by Pulse Field Gel Electrophoresis (PFGE) analysis (see Figure S1 and Supplemental Material) and one circular chromosome with an estimated size of 2,516 kb, with a

52% GC content. A total of 3,182 genes were predicted (3,049 protein-coding genes and 82 noncoding RNA genes (95.8% and 2.57% respectively)). *E. chiriqhucha* str. N139 has 10 ribosomal rRNA operons, confirmed by PFGE (see Supplemental Material and Figure S2). A putative function was assigned to 2,214 (73%) of the protein-coding genes, and the remaining genes were annotated as hypothetical proteins. The properties and the statistics of the genome are summarized in Table 3. 2,575 protein-coding genes were assigned to 1,603 COG families, corresponding to a gene content redundancy of 38.1% (see Table 4).

Genome Rearrangements

Genome rearrangements within clades I and II are scarce, showing high conservation of the genomic structure within clades. However, several genomic rearrangements occurred as both clades diverged.

In order to determine which contigs of *E. chiriqhucha* str. N139 belong to plasmids, the plasmid sequences of pEspA and pEspB from *E. arabatum* RFL1109 (Jakubauskas et al., 2009) were retrieved from NCBI. This strain was selected for comparison because their plasmids have been widely studied (Jakubauskas et al., 2009) and because it is phylogenetically close to *E. chiriqhucha* str. N139. Jakubauskas and colleagues identified the regions hr-A1, hr-AB and hr-A2 in plasmid pEspA as capable of replicating the plasmid in *Bacillus*. For plasmid pEspB, they hypothesized that the regions hr-B1, hr-AB and hr-B2 are involved in a theta replication mechanism (Jakubauskas et al., 2009). BLAST searches of these regions were performed against all *Exiguobacterium* genome sequences available to date. For the strains *E.* MH3, *E. antarcticum* and *E.* sp. AT1b, which are described as genomes without plasmids, no significant hits were found. Conversely, hits to the *E. arabatum* sequences hr-B1, hr-AB and hr-B2 (a fragment of 39 kb) were

246 found in the genomes of *E.* GIC31 (56 kb) and *E.* N139 (contig000014 of size 25 kb). It was
247 concluded that the sequences present in the plasmids are shared within different *Exiguobacterium*
248 strains, displaying a highly dynamic behavior. Therefore it was not possible to determine which
249 of our contigs correspond to the three megaplasmids observed in the PFGE experiments (see
250 Supplemental Material). Furthermore, contig 14 in our assembly corresponds to the smallest contig
251 of *E.* GIC31, so it could be a plasmid in both *Exiguobacterium* strains. Genes belonging to contig
252 14 are mostly hypothetical proteins, only 11 genes could be annotated. Of these, 9 correspond to
253 genes involved in mobile elements (antirestriction proteins, integrases and transposases), conjugal
254 transfer proteins and competence factors; one antibiotic resistance gene and a RNaseH. However,
255 contig 14 lies adjacent to contig 13, both accounting for a total size of 100 kb when synteny was
256 evaluated against *E. pavillionensis* RW-2. It is worth mentioning that contig 13 possesses most of
257 the genes responsible for metals resistance, but this region appears to be integrated in the
258 chromosome of *E.* GIC31. This highly dynamic behavior across strains, along with the presence
259 of several genes involved in mobility, suggests that, if both contigs belong to a plasmid, it might
260 be an integrative one. A MAUVE analysis performed between the strains N139, GIC31 and *E.*
261 *pavillionensis* RW-2 shows high synteny across their chromosomes. This idea that contigs 12, 13
262 and 14 might belong to the plasmids is supported by their shifts in GC skew (Figure 3). To all
263 appearances, the chromosomes within each of the two main clades of the *Exiguobacterium* species
264 are very similar, but quite distinct when compared between these clades (Figure 1).

265 **The *Exiguobacterium* strain N139 belongs to the *chiriqhucha* species along with *E.***
266 ***pavilionensis* str. RW-2 and *Exiguobacterium* sp. GIC31**

267 A phylogenomic reconstruction (Figure 1A) placed the strain N139 as most similar to
268 *Exiguobacterium* str. GIC31 (Vishnivetskaya et al., 2014) as well as to *E. pavilionensis* str. RW-2

(White III, Grassa & Suttle, 2013). ANI and AAI calculations of all clade II *Exiguobacterium* strains were performed and compared to our N139 strain, suggesting that *E. pavilionensis* str. RW-2, *Exiguobacterium* sp. GIC31 and this N139 strain belong to the same species since they share ANI values above 97% (Table S2) (Goris et al., 2007). Typically, the ANI values between genomes of the same species are above 95% (e.g., *E. coli*). ANI and AAI scores of all pairwise comparisons of the three proposed *Exiguobacterium chiriqhucha* strains exceed the 97% threshold (data not shown). Also relying on the ANI and AAI calculations, it was concluded that the outgroup of the *E. chiriqhucha* species could be *E. mexicanum*.

Exiguobacterium clade II pangenomes

To further understand the genomic properties of *E. chiriqhucha* str. N139 and its taxonomic positioning, we built the pangenome of nine *Exiguobacterium* strains from clade II, whose complete genomes were available at the time of analysis. This pangenome is composed of 5,267 genes; 2,116 of them belonging to the core genome and 1,664 being Strain Specific Genes (SSGs). The resulting pangenome shows a very conserved and cohesive pool of genes, despite their evolutionary distance and their remote geographic locations. Over two thousand genes compose the core genome, which represents a large core genome when compared to other pangenomes , and taking into account that the average genome size of *Exiguobacterium* strains, which is approximately three thousand genes. The SSGs are represented in a heatmap on Figure S3 where the clusterization of the *Exiguobacterium* strains is based on the presence (and abundance) or absence of their COG assignation. *Exiguobacterium* sp. S17 and *E. mexicanum* are exceptional for the fact that they possess a large pool of SSGs (Table S3). We speculate that some of these SSGs could have been acquired by Horizontal Gene Transfer (HGT) and retained to adapt to these

diverse environments, or equally likely, lost in some of the living taxa, due to lack of selective pressure in their respective niches.

Fifty-nine of the SSGs found in the *E. chiriqhucha* str. N139 were mapped on its genome to see if their distribution followed some bias (Figure 3). Throughout the contigs that are putative chromosomal regions, the SSGs appear to be randomly distributed. However, some of the SSGs are concentrated in the contigs 12 and 13, supporting the previous idea that these contigs may be part of the megaplasmids seen in the PFGE analysis (Figure S1). COGs were assigned to the SSGs as previously described for the *E. chiriqhucha* str. N139 genome. For the whole set of SSGs of the pangenome, COGs were successfully assigned to 66% of the genes, and are represented in a heatmap (Figure S3). However, most of the N139 SSGs could not be assigned to COGs, and for those that were successfully classified, the vast majority falls in the S and R (Poorly Characterized) COG categories, leaving open questions on which may be the unique strategies that N139 employs to adapt to the particular environment of Laguna Negra.

Main Metabolic Pathways, Amino acids, Nucleotides and Cofactors

Based on its genomic content, *E. chiriqhucha* str. N139 is probably a chemoheterotroph since it has two copies of *aioB* arsenite oxidase, which means it could obtain energy from arsenite oxidation. It has the complete pathway for glycolysis and it could synthesize acetyl-CoA, succinyl-CoA and isobutanoyl-CoA. It is a heterolactic fermenter, being able to produce lactate from pyruvate and ethanol from acetaldehyde. It has a complete TCA cycle, and it lacks the first two steps of the pentose phosphate pathway, but the rest of the pathway is present. Hence, its central metabolism is similar to *B. subtilis* (Blencke et al., 2003, 2006), but *E. chiriqhucha* str. N139 can synthesize more fermentation products, namely ethanol and formate. *E. chiriqhucha* str. N139 lacks the routes for synthesizing *de novo* phenylalanine and tyrosine, as well as the Branched Chain

Amino Acids (BCAA). However, it can synthesize tyrosine from phenylalanine, since it has the phenylalanine-4-hydroxylase regulated by a Tyr (UAC codon) T box riboswitch. Despite lacking the complete pathways for BCAA biosynthesis, it preserves the *ilvE* gene, a BCAA aminotransferase, which could probably synthesize any of the three BCAAs from available precursors. An interesting note on its tryptophan biosynthesis is that its biosynthetic operon is split in two transcription units: *trpEG* and *trpDCFBA*, which are separated in the chromosome, but co-regulated by a Trp T box riboswitch. Although this regulation is common in Firmicutes (Gutierrez-Preciado et al., 2005; Gutiérrez-Preciado, Yanofsky & Merino, 2007), the genome context of the *trp* operon is not, and it is interesting that this separation takes place at the synthesis of anthranilate. Moreover, the *trpEG* genes are regulated by a single T box, whilst the *trpDCFBA* operon is regulated by two T boxes in tandem. This could either mean that the separation of the pathway is a recent event and the regulation is being settled in order to coordinate both transcriptional units; or that this strain requires anthranilate (the product of *trpEG*) for something else. Certainly, one possibility is that *E. chiriqhucha* str. N139 exports anthranilate for a synthrophy with a partner(s) and the subsequent steps of the tryptophan biosynthetic pathway require a stricter regulation in order for the genes *trpDCFBA* to be expressed. Since *E. chiriqhucha* str. N139 lacks the biosynthetic pathways for five amino acids, a likely scenario is that this bacterium is sharing metabolites with other partners in Laguna Negra. This is supported by the observation that it is able to form part of a biofilm, and that in all of the amino acids tested it can only grow on serine and asparagine as a sole carbon source (see Table S1). Based on the metabolite tracer from Pathway Tools, it can be inferred that *E. chiriqhucha* str. N139 could synthesize phenylalanine as well as valine from serine or asparagine. In the same fashion, it cannot grow with phenylalanine as the sole carbon source. Therefore, the configuration of the genes involved in amino acid

337 metabolism might represent a requirement of amino acid syntrophy that needs further exploration
338 and testing. A second possibility is that *E. chiriqhucha* str. N139 is utilizing anthranilate for some
339 novel pathway. Anthranilate cannot be in excess with respect to tryptophan, since its excess could
340 decrease the availability of phosphoribosyl pyrophosphate (PRPP) for histidine synthesis (and
341 other reactions) (Merino, Jensen & Yanofsky, 2008). This novel pathway could be involved in
342 different functions that require either tryptophan or anthranilate as intermediates. Examples of
343 these functions are quorum sensing molecules in *Pseudomonas aeruginosa (Farrow & Pesci,*
344 *2007)*, plant hormones in *Azospirillum brasilense (Ge, Xie & Chen, 2006),* violacein in
345 *Chromobacterium violacein (Antônio & Creczynski-Pasa, 2004)* or antibiotics as in *Streptomyces*
346 *coelicolor (Amir-Heidari, Thirlway & Micklefield, 2008)*.

347 The regulation of biosynthetic and transporter genes through riboswitches is common in
348 Firmicutes, specially the members of Bacilli class. It has also been observed that transport and
349 biosynthesis of the same metabolite tend to be part of a regulon mediated by *in cis* elements, like
350 riboswitches (Gutiérrez-Preciado et al., 2009). Methionine can be synthesized and imported
351 through several strategies. Several SAM riboswitch regulated operons coding for Met transporters
352 were identified in the genome of *E. chiriqhucha* str. N139, as well as canonical *met* biosynthetic
353 genes. An interesting case is the methionine salvage pathway, whose genes are encoded in two
354 divergent operons, both regulated by divergent SAM riboswitches. Both operons must be
355 transcribed in order for the Yang cycle to be completed. In one operon, genes *mtnK* and *mtnA* are
356 transcribed along with three ribose transporters, *rbsB, rbsC* and *araG*. Lysine biosynthesis (from
357 aspartate via diaminopimelate) and transport are part of a regulon under the lysine riboswitch.
358 Furthermore, through the identification of riboswitches, two transporters from the NhaC family
359 were annotated: one as a methionine transporter (SAM riboswitch), and the other one as a lysine

transporter (LYS riboswitch). This strategy of improving gene annotation through the knowledge of the gene's regulation has been previously explored (Rollins, 2002; Gutiérrez-Preciado & Merino, 2012; Gutiérrez-Preciado et al., 2015).

Cofactors. Thiamine can be synthesized *de novo*, its biosynthesis and its uptake are regulated by the TPP riboswitch. Moreover, the analysis of *E. chiriqhucha* str. N139 genome indicates that a new thiamine transporter could be present in this bacterium. The gene exiN139_02072 is automatically identified as a membrane protein, but it seems to be regulated by a TPP riboswitch. Experimental evidence is needed for the confirmation and characterization of this transporter, which could unveil a new family of thiamine transporters. Riboflavin biosynthesis and transport (RibU) are also co-regulated through a FMN riboswitch.

Nucleotides. In the genome of *E. chiriqhucha* str. N139 the purine *de novo* biosynthetic pathway is encoded in a huge transcription unit regulated by a purine riboswitch. Other transcription units in the same regulon include a monocystronic GMP synthase, and genes involved in adenine and adenosine salvage pathway.

Genomic Adaptations to an Extreme Environment

Laguna Negra is an aquatic ecosystem that harbors extreme environmental conditions such as high levels of UV-B (10.65 wm2), high salinity levels (32%), scarce nutrients, particularly phosphorous (<005 mg/l), high metal contents including the metalloid arsenic (3mg/l), an alkaline pH and large daily temperature fluctuations (ranging from 20 °C during the day to −40 °C at night) (Flores et al., 2009); (see Table 1).

Resistance to metals and metalloids

In Laguna Negra, ubiquitous Arsenic enters the *E. chiriqhucha* str. N139 cells through existing transporters due to its high structural similarity with other molecules (Rosen, 1999) and induces oxidative stress responses (Oremland & Stolz, 2003). Furthermore, arsenite (AsO_2H) and arsenate (AsO_4^{3-}), are both toxic molecules. Arsenite binds to reduced cysteines in proteins inactivating them, and arsenate is a molecular analog of phosphate and therefore inhibits oxidative phosphorylation (Oremland & Stolz, 2003). Arsenate is far less toxic than arsenite, hence the oxidation of arsenite is considered a detoxification process. However, the oxidation of arsenite to arsenate, when coupled to the reduction of oxygen to water, is an exergonic process, and it has been suggested that at least some bacteria may derive energy out of this process (vanden Hoven & Santini, 2004). *E. chiriqhucha* str. N139 has an arsenite oxidase, AioB, enabling it to oxidize arsenite. This is an important metabolic capability, because it uses arsenite as an electron donor. Moreover, from a bioremediation point of view, this former metabolic feature is important since arsenite is more soluble than arsenate, so it can facilitate the removal of As in solution. *E. chiriqhucha* str. N139 also has an arsenite efflux pump, ArsB, as well as an ATPase that provides energy to ArsB for extrusion of arsenite and antimonite, ArsA, co-transcribed with ArsD, an arsenic chaperone for the ArsAB pump (Páez-Espino et al., 2009). Hence, this bacterium can probably detoxify and extrude As, as well as oxidize arsenite acquiring energy from this process. These ArsAB and ArsD proteins are also present in *Salinivibrio* strains isolated from the Laguna Socompa. However, these *Salinivibrio* strains also have ArsC, a cytoplasmic oxidoreductase that reduces arsenate to arsenite in a ATP-glutathione-glutaredoxin dependent way (Gorriti et al., 2014). *E. chiriqhucha* str. N139 lacks significant homologs to this gene as well as significant homologs to *B. subtilis' arsC* gene.

403 *E. chiriqhucha* str. N139 also possesses redundancy for mercury detoxification, harboring
404 four paralogous copies of *merA*. Briefly, MerA is the key detoxification enzyme of the mercury
405 resistance system, reducing Hg^{2+} to Hg^0 (Silver & Phung, 2005). Hg is toxic due to its high affinity
406 to sulfur (Nies, 2003) and usually, *mer* resistance genes are co-transcribed in an operon whose
407 dissemination is common by horizontal gene transfer (HGT) (Barkay, Miller & Summers, 2003).
408 In this organism, two copies of *merA* are present in a monocystronic fashion; a third one is
409 transcribed with a hypothetical protein. A fourth copy is co-transcribed with *merR*, the regulatory
410 protein of the system. Two *mer* transporters which uptake Hg and *merP*, a transporter with a Sec-
411 type signal, which could import Hg as a neutral chloride or hydroxide and deliver it to the other
412 Mer transporters, which will finally transfer it to MerA.

413 The most common mechanism of resistance to metals consists of efflux pumps for
414 inorganic ions. However, As and Hg resistance mechanisms are unique in the sense that these
415 elements are reduced to lower their toxicity (Silver & Phung, 2005), instead of being exported. *E.*
416 *chiriqhucha* str. N139 is resistant to cadmium, zinc, cobalt, and copper by pumping it out from the
417 cell. It has two membrane embedded Cd^{2+} efflux pumps, one of which can also extrude zinc and
418 cobalt; two paralogous copies of *copA* and *copB*, two P-type ATPase systems for exporting copper,
419 and *cueR,* a sensing cytoplasmic Cu that protects periplasmic proteins from copper-induced
420 toxicity (Orell et al., 2010). *copB* is transcribed monocystronically, and each of the *copA* genes
421 form an operon co-transcribed with a copy of *copZ*, a copper chaperone, but one is co-transcribed
422 with a glutaredoxin, whilst the other is co-transcribed with *csoR*, a copper-sensitive operon
423 repressor.

424 Additionally, this microorganism lives in a low phosphorous environment, and relies on
425 strategies for phosphorous uptake, like the presence of high-affinity Pi transporters and its

426 regulation (*pstS, pstCAB, phoB, phoR, dedA* and *ptrA*) and genes for polyphosphate storage and
427 breakdown (*ppk* and *ppx*). Organisms that scavenge phosphate can sometimes uptake the
428 structurally similar arsenate ion, and hence also depend on arsenate detoxification mechanisms. It
429 is also able to thrive in the alkaline environment of Laguna Negra since its genome code for all the
430 typical antiporters present in alkaliphilic bacilli (*nhaC, nhaP,* the *mpr* operon*, yhaU, norM* and
431 *mleN*). These antiporters present also contribute to a moderate salinity resistance this could also
432 be related with the maintenance of metal resistance strategies in its genome, since it has been
433 shown that lowering the salinity can lead to enhanced sensitivity to cadmium, cobalt and copper

434

435 ## Resistance to UV radiation

436 Another extreme environmental condition in Laguna Negra is high ultraviolet radiation,
437 particularly UV-B (Flores et al., 2009; Ordoñez et al., 2009). In order to determine if *E.*
438 *chiriqhucha* str. N139 can cope with this constant stress, we measure the effect of colony survival
439 of different *Exiguobacterium* strains exposed to UV-B radiation (Figure 4). More than 25% of the
440 colonies survive after 3 hrs of constant UV-B radiation. This contrasts with the other
441 *Exiguobacterium* strains which rapidly start to decay, even though one of them, str. S17 was
442 isolated from a neighbor lake, Laguna Socompa, in the HAALs (Ordoñez et al., 2013).

443 Bacteria have different UV damage repair pathways, including photoenzymatic repair
444 (PER), nucleotide excision repair (NER) also called dark repair, and recombinational repair (PRR)
445 or post-replication repair (Goosen & Moolenaar, 2008). *E. chiriqhucha* str. N139 has three genes
446 (exiN139_00335 (*phrB*), exiN139_01768 and exiN139_00235) related to photolyases, which are
447 involved in PER. They use UV as energy source (using FADH and transferring electrons) and

448 catalyze the monomerization of cyclobutyl pyrimidine dimers. The gene exiN139_00335 only has
449 homologues in Firmicutes including other known *Exiguobacterium,* and exiN139_01768 has
450 homologues in Firmicutes, Cyanobacteria, α- and γ- Proteobacteria, and Euryarcheotes.
451 Additionally, exiN139_00235 is a cryptochrome, which are flavoproteins related to photolyases.
452 Cryptochromes do not repair DNA and are presumed to act in other (unknown) processes, such as
453 entraining circadian rhythms (Yuan et al., 2012). It is worth to remark that from these proteins
454 only exiN139_01768, annotated as Deoxyribodipyrimidine photo-lyase-related protein, contains
455 significant homologs in the genomes of the two *Exiguobacterium* compared in the UV-B radiation
456 resistant assay (Figure 4).

457 *E. chiriqhucha* str. N139 has also genes for NER. Its genome encodes the UvrABC
458 endonuclease, a complex that recognizes DNA damage, binds to the damaged segment and cleaves
459 it. Additionally, it codes for PcrA (also known as UvrD), a helicase in charge of removing the
460 excised segment recognized and cleaved by UvrABC. These genes are regulated by the SOS
461 response, which uses LexA as a repressor inactivated by RecA (Minko et al., 2001). *E.*
462 *aurantiacum* and *E.* str. S17, contain homologs for the *uvrB/uvrC* gene, and the *pcrA* and *recA*
463 regulatory genes.

464 Regarding the PRR, *E. chiriqhucha* str. N139 encodes for RecA, which recognizes SSB
465 and cleaves UmuD, which becomes UmuD' and binds UmuC to generate polymerase V, which in
466 turn repairs damages, sometimes causing mutations. The gene exiN139_03003 may produce
467 polymerase IV that is also involved in DNA damage repair (Sommer, Bailone & Universitaire,
468 1998). UmuC is found in the genome, however UmuD is missing. It is possible that a protein
469 highly similar to an existing copy of LexA may be taking its role, given that both can be cleaved

by RecA and are present in *Exiguobacterium*. UmuD and UmuC are not present in the genomic sequences of *E. aurantiacum* and *E.* str. S17.

E. chiriqhucha str. N139 appears robust towards UV-B radiation, possessing several mechanism to cope with this constant stress. Moreover, the two strains proposed to be part of the same *chiriqhucha* species, *E. pavilionensis* str. RW-2 and *E.* str. GIC31, contain the same set of genes described above to cope with UV-B radiation, another unifying property of these strains. The strains RW-2 and N139 contain two homologs for Deoxyribodipyrimidine photo-lyases, while GIC31 contains three.

Living in syntrophy?

Finally, it is likely that *E. chiriqhucha* str. N139 participates in biofilm formation in Laguna Negra along with other bacteria (unpublished results). Analyses of other *Exiguobacterium* have shown that they participate in marine biofilms interacting with other Firmicutes and Proteobacteria (López et al., 2006; Carneiro et al., 2012). Evidence of possible biofilms associated genes originates from two *loci* present in the genome of *E. chiriqhucha* str. N139. The first *locus* encoding a protein capable of producing alginate, a linear co-polymer of two uronic acids that is produced in its acetylated form by some bacteria for adherence of these bacteria to target cell walls by the creation of a biofilm (Ramphall & Pier, 1985). The second *locus* codes for the arginine deiminase system, which can function at very low pH and is thought to be a critical factor in oral biofilm pH homeostasis (Burne & Marquis, 2000).

Conclusions

E. chiriqhucha str. N139 lives in a high-altitude, salted lake exposed to intense UV radiation, about 300 km away from the nearest ocean, the Pacific. Many factors in *E. chiriqhucha*

492 str. N139 metabolism, such as the its needs to uptake certain intermediates like phenylalanine and
493 BCAAs, and the possible excretion of the overproduced anthranilate, suggest that it is a key player
494 in the amino acid metabolism of a microbial consortium that inhabit Laguna Negra. Moreover, the
495 excess of anthranilate that it may produce could be directed to some novel pathway that remains
496 to be uncovered, such as a new antibiotic, a new pigment or a new quorum sensing molecule.

497 The genome of *E. chiriqhucha* str. N139 contains all the necessary strategies to cope with
498 all the environmental stresses that simultaneously co-occur in Laguna Negra. This
499 *Exiguobacterium* is able to detoxify metals like arsenic, mercury, cadmium, zinc, cobalt, and
500 copper; it has a complete defense system against UV damage; and it is also able to thrive in the
501 alkaline environment of Laguna Negra. (Ventosa, Nieto & Oren, 1998). With all these
502 characteristics, *E. chiriqhucha* str. N139 is an excellent candidate for future biotechnological
503 research.

504

505 Although our study generates more questions than the ones it could solve, by sequencing
506 its genome we have gained insights on the strategies the strain N139 employs for thriving in its
507 habitat. From its Strain Specific set of genes, only 23 out of 59 could be annotated and classified
508 to a COG, and still, most of the COG-classified genes belong to the poorly characterized category.
509 This set of genes of unknown function require further experimental work to completely unveil how
510 the strain N139 is adapting to the extreme environment of Laguna Negra.

511

512

Description of *Exiguobacterium chiriqhucha sp. nov.*

Exiguobacterium chiriqhucha (chi.ri.qhu.cha. (/ʃi ri ku tʃɑ /) Quechua. Adj. *chiri*: cold, freezing; Quechua. Noun. *qhucha:* lake, pond. *chiriqhucha* of or belonging to a cold lake, referring to the common habitat of these three species). Members of the species *Exiguobacterium chiriqhucha*; inhabitants of freshwater ponds, saline ponds; distinguishable by their 16S rRNA sequences; accession numbers are: JMEH00000000 for the str. N139 genome, ATCL00000000 for the RW-2 genome (White III, Grassa & Suttle, 2013) and JNIP00000000 for the GIC31 genome (Vishnivetskaya et al., 2014). The three strains that so far comprise this species form orange shiny colonies and are Gram-positive, rod-shaped, facultative anaerobes and motile via peritrichous flagella (Miteva, Sheridan & Brenchley, 2004; Vishnivetskaya, Kathariou & Tiedje, 2009; White III, Grassa & Suttle, 2013). Two of them, str. RW-2 and str. GIC31 were isolated from permanently cold environments (Pavilion Lake and Glacier Ice, Greenland;(White III, Grassa & Suttle, 2013; Vishnivetskaya et al., 2014)) whilst the str. N139 was isolated from Laguna Negra, a HAAL which temperature can drop to -30ºC. Their temperature range of growth is from a minimum (str. GIC31) of 2ºC to a maximum (str. RW-2) of 50ºC; the pH range for growth is from 5 to 11 in str. RW-2 and from 7 to 9 in str. N139 (Miteva, Sheridan & Brenchley, 2004; White III, Grassa & Suttle, 2013). The three strains possess cold-shock proteins and have a G+C content of 52% (White III, Grassa & Suttle, 2013; Vishnivetskaya et al., 2014). The type strain is RW-2.

Acknowledgements

We thank Jerome Verleyen for computer support and the Instituto de Biotecnología-UNAM for allowing us access to its computer cluster. We would also like to thank Eric Nawrocki for his guidance with the *ssu-align* program.

Tables

Table 1. Classification and general features of *Exiguobacterium chiriqhucha* str. N139

Property	Term	Evidence code[a]
Classification	Domain *Bacteria*	TAS (Woese, Kandler & Wheelis, 1990)
	Phylum Firmicutes	TAS (Gibbons & Murray, 1978)
	Class Bacilli	TAS (De Vos et al., 2009)
	Order Bacillales	TAS (De Vos et al., 2009)
	Family Bacillales *Family XII. Incertae Sedis*	TAS (De Vos et al., 2009)
	Genus Exiguobacterium	TAS (De Vos et al., 2009; Vishnivetskaya, Kathariou & Tiedje, 2009)
	Species *Exiguobacterium chiriqhucha*	TAS (White III, Grassa & Suttle, 2013)
	Strain: *N139 (Accession: JMEH00000000)*	
Gram stain	*Positive*	IDA
Cell shape	*Short rods*	IDA
Motility	*Motile*	IDA
Sporulation	*Non-sporulating*	EXP
Temperature range	*Mesophilic (30 - 37°C)*	IDA
Optimum temperature	*30°C*	IDA
pH range; Optimum	*7-9*	IDA
Carbon source	*β-Methylglucoside, Galacturonic acid, L-asparagine, Tween 40, L-Serine, N-acetyl-glucosamine, Hydroxybutyric acid, Itaconic acid, Ketobutyric acid, Putrescine (See table S1)*	EXP
Habitat	*Aquatic*	TAS (Flores et al., 2009)
Salinity	*0.11% - 10% NaCl (w/v)*	IDA
Oxygen requirement	*Facultatively anaerobic*	TAS (De Vos et al., 2009)
Biotic relationship	*free-living*	IDA
Pathogenicity	*non-pathogen*	NAS
Geographic location	*Laguna Negra, Catamarca, Argentina*	IDA
Sample collection	*2006*	IDA

Latitude	*27°39'20.17"S*	IDA
Longitude	*68°33'46.18"W*	IDA
Altitude	*4100 masl*	IDA

[a] Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence); EXP: Inferred from Experiment. These evidence codes are from the Gene Ontology project ("Gene Ontology Evidence Codes")

Table 2. Project information.

Property	Term
Finishing quality	Permanent-draft
Libraries used	454 pyrosequence standard library
Sequencing platforms	454 Titanium
Fold coverage	85 x
Assemblers	Newbler 2.8 and MIRA 3.4
Gene calling method	Prokka
Locus Tag	EF88
Genbank ID	JMEH00000000.1
GenBank Date of Release	December, 2015
GOLD ID	Go0093977
BIOPROJECT	PRJNA245187
Source Material Identifier	N139
Project Relevance	UV resistance, metal resistance, adaptation to oligotrophic environments

Table 3. Nucleotide content and gene count levels of the *E. chiriqhucha* str. N139 genome

Attribute	Genome (total)	
	Value	% of total[a]
Genome size (bp)	2,952,588	-
DNA coding (bp)	2,655,834	89.94
DNA G+C (bp)		52
DNA Scaffolds	23	
N50	1,553,709	
Total genes	3,182	100
RNA genes	82	2.62
Protein-coding genes	3,049	95.82
Pseudogenes	26	0.81
Genes in internal clusters	NA	

Genes with function prediction	2,356	74.04
Genes assigned to COGs	2,575	80.92
Genes with Pfam domains	2,538	79.76
Genes with signal peptides	NA	
Genes with transmembran helices	888	27.90
CRISPR repeats	0	

a) The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.

Table 4. Genes associated with the 25 general COG functional categories

Code	Value	% of total[a]	Description
J	166	5.44	Translation, ribosomal structure and biogenesis
A	0	0	RNA processing and modification
K	235	7.71	Transcription
L	144	4.72	Replication, recombination and repair
B	1	0.03	Chromatin structure and dynamics
D	36	1.18	Cell cycle control, Cell division, chromosome partitioning
V	62	2.03	Defense mechanisms
T	166	5.44	Signal transduction mechanisms
M	144	4.72	Cell wall/membrane biogenesis
N	75	2.46	Cell motility
U	53	1.74	Intracellular trafficking and secretion
O	100	3.28	Posttranslational modification, protein turnover, chaperones
C	152	4.99	Energy production and conversion
G	232	7.61	Carbohydrate transport and metabolism
E	224	7.35	Amino acid transport and metabolism
F	84	2.76	Nucleotide transport and metabolism
H	97	3.18	Coenzyme transport and metabolism
I	81	2.66	Lipid transport and metabolism
P	170	5.58	Inorganic ion transport and metabolism
Q	54	1.77	Secondary metabolites biosynthesis, transport and catabolism
R	463	15.19	General function prediction only
S	327	10.72	Function unknown
-	447	15.55	Not in COG

[a]The total is based on the total number of protein coding genes in the annotated genome.

Figure Legends.

Figure 1. **Evolutionary history of the genus** *Exiguobacterium*. A) Phylogenetic reconstruction using complete genomic sequences of 17 representative *Exiguobacterium* strains. The tree was built with PhyloPhlAn (Segata et al., 2013). B) **Synteny among *Exiguobacterium* strains**. Nucleotide syntenic blocks are represented by colored bars. Red links denote no rearrangements between the blocks compared. Blue links denote rearrangements between the blocks compared. Blue numbers in the phylogeny denote the minimum number of rearrangements obtained with *MGR*. Plasmids from *E. sibiricum* are displayed at the right (separated from the chromosome by backslashes). Black numbers indicate bootstrap values different from 100%.

Figure 2. **Differential interference contrast image of *E. chiriqhucha* str. N139**.

Figure 3. **Circular genome map of *E. chiriqhucha* str. N139**. Circle tracks from out towards inside are as follows: 1) Length in nucleotides for each contig; 2) Coding Sequences (CDS) in the Forward Strand (light blue); 3) CDS in the reverse strand (dark blue); 4) Strain Specific Genes (SSGs) in the forward strand (light purple); 5) SSGs in the reverse strand (dark purple); 6) GC Skew (gray). Skew and gene distribution follow that of a typical Firmicute genome. The Strain Specific Genes in the contigs that belong to the chromosome appear to be randomly distributed, whilst they seem to be concentrated in the contigs 12 and 13, which are probably the ones belonging to megaplasmids. The circular plot was done with Circos software (Krzywinski et al., 2009).

Figure 4. **Effect of ultraviolet B (UV-B) radiation on Exiguobacterium strains.** Percentage survival to UV-B radiation of str. N139 (dark circle), str. S17 (light circle) and str. DSMZ 6208 (dark triangle). The influence of UV-B radiation was studied by exposing liquid cultures to increasing doses, varying exposure times between 0 and 240 minutes.

583

584

585 ## References

586

587 Albarracín VH., Dib JR., Ordoñez OF., Farías ME. 2011. A Harsh Life To Indigenous
588 Proteobacteria At the Andean Mountains: Microbial Diversity and Resistance. In: Sezenna
589 ML ed. *Proteobacteria: Phylogeny, Metabolic Diversity and Ecological Effects*.

590 Amir-Heidari B., Thirlway J., Micklefield J. 2008. Auxotrophic-precursor directed biosynthesis
591 of nonribosomal lipopeptides with modified tryptophan residues. *Organic & biomolecular*
592 *chemistry* 6:975–978. DOI: 10.1039/b718766c.

593 Antônio RV., Creczynski-Pasa TB. 2004. Genetic analysis of violacein biosynthesis by
594 Chromobacterium violaceum. *Genetics and Molecular Research* 3:85–91.

595 Barkay T., Miller SM., Summers AO. 2003. Bacterial mercury resistance from atoms to
596 ecosystems. *FEMS Microbiology Reviews* 27:355–384. DOI: 10.1016/S0168-
597 6445(03)00046-9.

598 Belfiore C., Ordoñez OF., Farías ME. 2013. Proteomic approach of adaptive response to arsenic
599 stress in Exiguobacterium sp. S17, an extremophile strain isolated from a high-altitude
600 Andean Lake stromatolite. *Extremophiles : life under extreme conditions* 17:421–31. DOI:
601 10.1007/s00792-013-0523-y.

602 Blencke H-M., Homuth G., Ludwig H., Mäder U., Hecker M., Stülke J. 2003. Transcriptional
603 profiling of gene expression in response to glucose in Bacillus subtilis: regulation of the
604 central metabolic pathways. *Metabolic Engineering* 5:133–149. DOI: 10.1016/S1096-
605 7176(03)00009-0.

606 Blencke H-M., Reif I., Commichau FM., Detsch C., Wacker I., Ludwig H., Stülke J. 2006.
607 Regulation of citB expression in Bacillus subtilis: integration of multiple metabolic signals
608 in the citrate pool and by the general nitrogen regulatory system. *Arch Microbiol* 185:136–
609 146. DOI: 10.1007/s00203-005-0078-0.

610 Bourque G., Pevzner P a. 2002. Genome-scale evolution: Reconstructing gene orders in the
611 ancestral species. *Genome Research* 12:26–36.

612 Burge SW., Daub J., Eberhardt R., Tate J., Barquist L., Nawrocki EP., Eddy SR., Gardner PP.,
613 Bateman A. 2013. Rfam 11.0: 10 years of RNA families. *Nucleic acids research* 41:D226-
614 32. DOI: 10.1093/nar/gks1005.

615 Burne R a., Marquis RE. 2000. Alkali production by oral bacteria and protection against dental
616 caries. *FEMS Microbiology Letters* 193:1–6. DOI: 10.1111/j.1574-6968.2000.tb09393.x.

617 Cabrol N., McKay C., Grin E., Kiss K., Acs E., Toth B., Grigorszky I., Szabo K., Fike D., Hock
618 A., Demergasso C., Escudero L., Galleguillos P., Chong G., Grigsby B., Zambrana Roman
619 J., Tambley C. 2007. Signatures of habitats and life in Earth's high-altitude lakes: Clues to
620 Noachian aqueous environments on Mars. In: Chapman M ed. *The geology of Mars:*
621 *Evidence from Earth-based analogs*. Cambridge: Cambridge University Press, 349–370.

622 Carneiro AR., Ramos RTJ., Dall'Agnol H., Pinto AC., Soares SDC., Santos AR., Guimarães
623 LC., Almeida SS., Baraúna RA., das Graças DA., Franco LC., Ali A., Hassan SS., Nunes
624 CIP., Barbosa MS., Fiaux KK., Aburjaile FF., Barbosa EGV., Bakhtiar SM., Vilela D.,
625 Nóbrega F., dos Santos AL., Carepo MSP., Azevedo V., Schneider MPC., Pellizari VH.,
626 Silva A., de Castro Soares S., Santos AR., Guimarães LC., Almeida SS., Baraúna RA., das
627 Graças DA., Franco LC., Ali A., Hassan SS., Nunes CIP., Barbosa MS., Fiaux KK.,
628 Aburjaile FF., Barbosa EGV., Bakhtiar SM., Vilela D., Nóbrega F., dos Santos AL., Carepo
629 MSP., Azevedo V., Schneider MPC., Pellizari VH., Silva A. 2012. Genome sequence of
630 Exiguobacterium antarcticum B7, isolated from a biofilm in Ginger Lake, King George
631 Island, Antarctica. *Journal of Bacteriology* 194:6689–90. DOI: 10.1128/JB.01791-12.

632 Caspi R., Altman T., Billington R., Dreher K., Foerster H., Fulcher C a., Holland T a., Keseler
633 IM., Kothari A., Kubo A., Krummenacker M., Latendresse M., Mueller L a., Ong Q., Paley
634 S., Subhraveti P., Weaver DS., Weerasinghe D., Zhang P., Karp PD. 2014. The MetaCyc
635 database of metabolic pathways and enzymes and the BioCyc collection of
636 Pathway/Genome Databases. *Nucleic Acids Research* 42:459–471. DOI:
637 10.1093/nar/gkt1103.

638 Chevreux B., Pfisterer T., Drescher B., Driesel AJ., Müller WEG., Wetter T., Suhai S. 2004.
639 Using the miraEST assembler for reliable and automated mRNA transcript assembly and
640 SNP detection in sequenced ESTs. *Genome Research* 14:1147–1159. DOI:
641 10.1101/gr.1917404.

642 Consortium TU. 2015. UniProt : a hub for protein information. 43:204–212. DOI:
643 10.1093/nar/gku989.

644 Crapart S., Fardeau ML., Cayol JL., Thomas P., Sery C., Ollivier B., Combet-Blanc Y. 2007.
645 Exiguobacterium profundum sp. nov., a moderately thermophilic, lactic acid-producing
646 bacterium isolated from a deep-sea hydrothermal vent. *International Journal of Systematic*
647 *and Evolutionary Microbiology* 57:287–292. DOI: 10.1099/ijs.0.64639-0.

648 Darling AC., Mau B., Blattner FR., Perna NT. 2004. Mauve: multiple alignment of conserved
649 genomic sequence with rearrangements. *Genome Res* 14:1394–1403.

650 Dib J., Motok J., Fernández-Zenoff V., Ordoñez O., Farías ME. 2008. Occurrence of resistance
651 to antibiotics, UV-B, and arsenic in bacteria isolated from extreme environments in high-
652 altitude (above 4400 m) Andean wetlands. *Current microbiology* 56:510–7. DOI:
653 10.1007/s00284-008-9103-2.

654 Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space
655 complexity. *BMC bioinformatics* 5:113. DOI: 10.1186/1471-2105-5-113.

656 Ericksen GE., Salas RA. 1987. *Geology and Resources of Salars in the Central Andes*. U.S.

Geological Survey.

Farías ME., Fernández-Zenoff V., Flores R., Ordóñez O., Estévez C. 2009. Impact of solar radiation on bacterioplankton in laguna vilama, a hypersaline andean lake (4650 m). *Journal of Geophysical Research: Biogeosciences* 114:1–8. DOI: 10.1029/2008JG000784.

Farrow JM., Pesci EC. 2007. Two distinct pathways supply anthranilate as a precursor of the Pseudomonas quinolone signal. *Journal of Bacteriology* 189:3425–3433. DOI: 10.1128/JB.00209-07.

Fernández-Zenoff V., Heredia J., Ferrero M., Siñeriz F., Farías ME. 2006. Diverse UV-B resistance of culturable bacterial community from high-altitude wetland water. *Current microbiology* 52:359–62. DOI: 10.1007/s00284-005-0241-5.

Fernández-Zenoff V., Siñeriz F., Farías ME. 2006. Diverse responses to UV-B radiation and repair mechanisms of bacteria isolated from high-altitude aquatic environments. *Applied and environmental microbiology* 72:7857–63. DOI: 10.1128/AEM.01333-06.

Fischer S., Brunk BP., Chen F., Gao X., Harb OS., Iodice JB., Shanmugam D., Roos DS., Stoeckert CJ. 2011. Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes Into New Ortholog Groups. *Current protocols in bioinformatics* Chapter 6:Unit6.12.

Flores MR., Ordoñez OF., Maldonado MJ., Farías ME. 2009. Isolation of UV-B resistant bacteria from two high altitude Andean lakes (4 , 400 m) with saline and non saline conditions. *J. Gen. Appl. Microbiol.* 55:447–458.

Ge SM., Xie BE., Chen SF. 2006. Characterization of two trpE genes encoding anthranilate synthase ??-subunit in Azospirillum brasilense. *Biochemical and Biophysical Research Communications* 341:494–499. DOI: 10.1016/j.bbrc.2006.01.009.

Gene Ontology Evidence Codes. *Available at http://geneontology.org/page/guide-go-evidence-codes*

Gibbons NE., Murray RGE. 1978. Proposals Concerning the Higher Taxa of Bacteria. *International Journal of Systematic Bacteriology* 28:1–6. DOI: 10.1099/00207713-28-1-1.

Goosen N., Moolenaar GF. 2008. Repair of UV damage in bacteria. *DNA Repair* 7:353–379. DOI: 10.1016/j.dnarep.2007.09.002.

Goris J., Konstantinidis KT., Klappenbach J a., Coenye T., Vandamme P., Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* 57:81–91. DOI: 10.1099/ijs.0.64483-0.

Gorriti MF., Dias GM., Chimetto LA., Trindade-silva AE., Silva BS., Mesquita MMA., Gregoracci GB., Farias ME., Thompson CC., Thompson FL. 2014. Genomic and phenotypic attributes of novel Salinivibrios from stromatolites , sediment and water from a high altitude lake. *BMC Genomics* 15:473.

694 Gutiérrez-Preciado A., Henkin TM., Grundy FJ., Yanofsky C., Merino E. 2009. Biochemical
695 features and functional implications of the RNA-based T-box regulatory mechanism.
696 *Microbiology and molecular biology reviews MMBR* 73:36–61. DOI:
697 10.1128/MMBR.00026-08.

698 Gutierrez-Preciado A., Jensen R a., Yanofsky C., Merino E. 2005. New insights into regulation
699 of the tryptophan biosynthetic operon in Gram-positive bacteria. *Trends in Genetics*
700 21:432–6. DOI: 10.1016/j.tig.2005.06.001.

701 Gutiérrez-Preciado A., Merino E. 2012. Elucidating metabolic pathways and digging for genes of
702 unknown function in microbial communities: the riboswitch approach. *Clinical*
703 *microbiology and infection : the official publication of the European Society of Clinical*
704 *Microbiology and Infectious Diseases* 18 Suppl 4:35–9. DOI: 10.1111/j.1469-
705 0691.2012.03864.x.

706 Gutiérrez-Preciado A., Torres AG., Merino E., Bonomi HR., Goldbaum FA., García-Angulo VA.
707 2015. Extensive Identification of Bacterial Riboflavin Transporters and Their Distribution
708 across Bacterial Species. *Plos One* 10:e0126124. DOI: 10.1371/journal.pone.0126124.

709 Gutiérrez-Preciado A., Yanofsky C., Merino E. 2007. Comparison of tryptophan biosynthetic
710 operon regulation in different Gram-positive bacterial species. *Trends in Genetics* 23:422–
711 426. DOI: 10.1016/j.tig.2007.05.005.

712 Guy L., Kultima JR., Andersson SGE. 2010. genoPlotR: Comparative gene and genome
713 visualization in R. *Bioinformatics* 26:2334–2335. DOI: 10.1093/bioinformatics/btq413.

714 vanden Hoven RN., Santini JM. 2004. Arsenite oxidation by the heterotroph Hydrogenophaga
715 sp. str. NT-14: the arsenite oxidase and its physiological electron acceptor. *Biochimica et*
716 *biophysica acta* 1656:148–55. DOI: 10.1016/j.bbabio.2004.03.001.

717 Jakubauskas A., Kriukiene E., Trinkunaite L., Sapranauskas R., Jurenaite-Urbanaviciene S.,
718 Lubys A. 2009. Bioinformatic and partial functional analysis of pEspA and pEspB, two
719 plasmids from Exiguobacterium arabatum sp. nov. RFL1109. *Plasmid* 61:52–64. DOI:
720 10.1016/j.plasmid.2008.09.004.

721 Jiang X., Xue Y., Wang L., Yu B., Ma Y. 2013. Genome Sequence of a Novel Polymer-Grade L
722 -Lactate-Producing Alkaliphile, Exiguobacterium sp. Strain 8-11-1. *Genome*
723 *Announcements* 1:4–5. DOI: 10.1016/j.biortech.

724 Karami K., Zolgharnei H., Assadi MM., Savari a., Dadollahi S. 2011. New Report on the
725 Occurrence of Exiguobacterium sp. AT1b in the Persian Gulf and its Resistance to Mercury
726 Pollution. *Current Research in Bacteriology* 4:23–27. DOI: 10.3923/crb.2011.23.27.

727 Karp PD., Paley SM., Krummenacker M., Latendresse M., Dale JM., Lee TJ., Kaipa P., Gilham
728 F., Spaulding A., Popescu L., Altman T., Paulsen I., Keseler IM., Caspi R. 2009. Pathway
729 Tools version 13.0: Integrated software for pathway/genome informatics and systems
730 biology. *Briefings in Bioinformatics* 11:40–79. DOI: 10.1093/bib/bbp043.

731 Kim IG., Lee MH., Jung SY., Song JJ., Oh TK., Yoon JH. 2005. Exiguobacterium aestuarii sp.

732 nov. and Exiguobacterium marinum sp. nov., isolated from a tidal flat of the Yellow Sea in
733 Korea. *International Journal of Systematic and Evolutionary Microbiology* 55:885–889.
734 DOI: 10.1099/ijs.0.63308-0.

735 Krzywinski M., Schein J., Birol I., Connors J., Gascoyne R., Horsman D., Jones SJ., Marra MA.
736 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–
737 1645.

738 Li L., Jr CJS., Roos DS. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic
739 Genomes. *Genome Research* 13:2178–2189. DOI: 10.1101/gr.1224503.candidates.

740 López-Cortés A., Schumann P., Pukall R., Stackebrandt E. 2006. Exiguobacterium mexicanum
741 sp. nov. and Exiguobacterium artemiae sp. nov., isolated from the brine shrimp Artemia
742 franciscana. *Systematic and Applied Microbiology* 29:183–190. DOI:
743 10.1016/j.syapm.2005.09.007.

744 López M a., Zavala-Díaz de la Serna FJ., Jan-Roblero J., Romero JM., Hernández-Rodríguez C.
745 2006. Phylogenetic analysis of a biofilm bacterial population in a water pipeline in the Gulf
746 of Mexico. *FEMS microbiology ecology* 58:145–54. DOI: 10.1111/j.1574-
747 6941.2006.00137.x.

748 Merino E., Jensen R a., Yanofsky C. 2008. Evolution of bacterial trp operons and their
749 regulation. *Current Opinion in Microbiology* 11:78–86. DOI: 10.1016/j.mib.2008.02.005.

750 Minko I., Hattman S., Lloyd RS., Kossykh V. 2001. Methylation by a mutant T2 DNA [N(6)-
751 adenine] methyltransferase expands the usage of RecA-assisted endonuclease (RARE)
752 cleavage. *Nucleic acids research* 29:1484–1490.

753 Mistry J., Finn RD., Eddy SR., Bateman A., Punta M. 2013. Challenges in homology search:
754 HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research* 41.
755 DOI: 10.1093/nar/gkt263.

756 Miteva VI., Sheridan PP., Brenchley JE. 2004. Phylogenetic and Physiological Diversity of
757 Microorganisms Isolated from a Deep Greenland Glacier Ice Core. *Appl Environ Microbiol*
758 70:202–213. DOI: 10.1128/AEM.70.1.202.

759 Nawrocki EP., Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches.
760 *Bioinformatics* 29:2933–2935. DOI: 10.1093/bioinformatics/btt509.

761 Nies DH. 2003. Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiology*
762 *Reviews* 27:313–339. DOI: 10.1016/S0168-6445(03)00048-2.

763 Ordoñez OF., Flores MR., Dib JR., Paz A., Farías ME. 2009. Extremophile culture collection
764 from Andean lakes: extreme pristine environments that host a wide diversity of
765 microorganisms with tolerance to UV radiation. *Microbial Ecology* 58:461–473. DOI:
766 10.1007/s00248-009-9527-7.

767 Ordoñez OF., Lanzarotti E., Kurth D., Gorriti MF., Revale S., Cortez N., Vazquez MP., Farias
768 ME., Turjanski AG., Farías ME., Turjanski AG. 2013. Draft Genome Sequence of the
769 Polyextremophilic Exiguobacterium sp . Strain S17, Isolated from Hyperarsenic Lakes in

the Argentinian Puna. *Genome Announcements* 1:e00480-13. DOI: 10.1186/1471-2164-9-75.Rodrigues.

Orell A., Navarro C a., Arancibia R., Mobarec JC., Jerez C a. 2010. Life in blue: copper resistance mechanisms of bacteria and archaea used in industrial biomining of minerals. *Biotechnology advances* 28:839–48. DOI: 10.1016/j.biotechadv.2010.07.003.

Oremland RS., Stolz JF. 2003. The ecology of arsenic. *Science (New York, N.Y.)* 300:939–44. DOI: 10.1126/science.1081903.

Páez-Espino D., Tamames J., de Lorenzo V., Cánovas D. 2009. Microbial responses to environmental arsenic. *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine* 22:117–30. DOI: 10.1007/s10534-008-9195-y.

Quevillon E., Silventoinen V., Pillai S., Harte N., Mulder N., Apweiler R., Lopez R. 2005. InterProScan: Protein domains identifier. *Nucleic Acids Research* 33:116–120. DOI: 10.1093/nar/gki442.

Ramphall R., Pier GB. 1985. Role of Pseudomonas aeruginosa Mucoid Exopolysaccharide in Adherence to Tracheal Cells. *Infection and Immunity* 47:1–4.

Rebollar E a., Avitia M., Eguiarte LE., González-González A., Mora L., Bonilla-Rosso G., Souza V. 2012. Water-sediment niche differentiation in ancient marine lineages of Exiguobacterium endemic to the Cuatro Cienegas Basin. *Environmental Microbiology* 14:2323–33. DOI: 10.1111/j.1462-2920.2012.02784.x.

Rissman AI., Mau B., Biehl BS., Darling AE., Glasner JD., Perna NT. 2009. Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics* 25:2071–2073. DOI: 10.1093/bioinformatics/btp356.

Rodrigues DF., Ivanova N., He Z., Huebner M., Zhou J., Tiedje JM. 2008. Architecture of thermal adaptation in an Exiguobacterium sibiricum strain isolated from 3 million year old permafrost: a genome and transcriptome approach. *BMC genomics* 9:547. DOI: 10.1186/1471-2164-9-547.

Rollins SM. 2002. The mRNA/tRNA interaction promoting T box transcriptional antitermination. The Ohio State University, Columbus, OH.

Rosen BP. 1999. Families of arsenic transporters. *Trends in microbiology* 7:207–12.

Rout SP., Rai A., Humphreys PN. 2015. Draft Genome Sequence of Alkaliphilic Exiguobacterium sp . Strain HUD , Isolated from a Polymicrobial Consortia. *Genome Announcements* 3:9–10. DOI: 10.1128/genomeA.01451-14.Copyright.

Sacheti P., Bhonsle H., Patil R., Kulkarni MJ., Srikanth R., Gade W. 2013. Arsenomics of Exiguobacterium sp. PS (NCIM 5463). *RSC Advances* 3:9705. DOI: 10.1039/c3ra40897c.

Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. DOI: 10.1093/bioinformatics/btu153.

Segata N., Börnigen D., Morgan XC., Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature communications* 4:2304. DOI: 10.1038/ncomms3304.

Silver S., Phung LT. 2005. A bacterial view of the periodic table: genes and proteins for toxic inorganic ions. *Journal of industrial microbiology & biotechnology* 32:587–605. DOI: 10.1007/s10295-005-0019-6.

Sommer S., Bailone A., Universitaire C. 1998. Specific RecA amino acid changes affect RecA – UmuDJC interaction. *Molecular Microbiology* 28:281–291.

Ventosa A., Nieto JJ., Oren A. 1998. Biology of Moderately Halophilic Aerobic Bacteria. *Microbiology and Molecular Biology Reviews* 62:504–544.

Vishnivetskaya TA., Chauhan A., Layton AC., Pfiffner SM., Huntemann M., Copeland A., Chen A., Kyrpides NC., Markowitz VM., Palaniappan K., Ivanova N., Mikhailova N., Ovchinnikova G., Andersen EW., Pati A., Stamatis D., Reddy TBK., Shapiro N., Nordberg HP., Cantor MN., Hua XS. 2014. Draft Genome Sequences of 10 Strains of the Genus Exiguobacterium. 2:10–11. DOI: 10.1128/genomeA.01058-14.Copyright.

Vishnivetskaya T a., Kathariou S. 2005. Putative transposases conserved in Exiguobacterium isolates from ancient Siberian permafrost and from contemporary surface habitats. *Applied and Environmental Microbiology* 71:6954–6962. DOI: 10.1128/AEM.71.11.6954-6962.2005.

Vishnivetskaya T a., Kathariou S., Tiedje JM. 2009. The Exiguobacterium genus: biodiversity and biogeography. *Extremophiles* 13:541–55. DOI: 10.1007/s00792-009-0243-5.

Vishnivetskaya T a., Lucas S., Copeland A., Lapidus A., del Rio TG., Dalin E., Tice H., Bruce DC., Goodwin L a., Pitluck S., Saunders E., Brettin T., Detter C., Han C., Larimer F., Land ML., Hauser LJ., Kyrpides NC., Ovchinnikova G., Kathariou S., Ramaley RF., Rodrigues DF., Hendrix C., Richardson P., Tiedje JM. 2011. Complete genome sequence of the thermophilic bacterium Exiguobacterium sp. AT1b. *Journal of Bacteriology* 193:2880–2881. DOI: 10.1128/JB.00303-11.

Vishnivetskaya T a., Petrova M a., Urbance J., Ponder M., Moyer CL., Gilichinsky D a., Tiedje JM. 2006. Bacterial community in ancient Siberian permafrost as characterized by culture and culture-independent methods. *Astrobiology* 6:400–414. DOI: 10.1089/ast.2006.6.400.

De Vos P., Garrity GM., Jones D., Krieg NR., Ludwig W., Rainey FA., Schleifer KH., Whitman WB. (eds.) 2009. *Bergey's Manual of Systematic Bacteriology*. New York: Springer.

White III RA., Grassa CJ., Suttle CA. 2013. Draft genome sequence of Exiguobacterium pavilionensis Strain RW-2, with Wide Thermal, Salinity, and pH Tolerance, Isolated from Modern Freshwater Microbioalites. *Genome announcements* 1:e0057-13. DOI: doi:10.1128/genomeA.00597-13.

Woese CR., Kandler O., Wheelis ML. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of*

Sciences of the United States of America 87:4576–4579. DOI: 10.1073/pnas.87.12.4576.

Yuan M., Chen M., Zhang W., Lu W., Wang J., Yang M., Zhao P., Tang R., Li X., Hao Y., Zhou Z., Zhan Y., Yu H., Teng C., Yan Y., Ping S., Wang Y., Lin M. 2012. Genome sequence and transcriptome analysis of the radioresistant bacterium deinococcus gobiensis: Insights into the extreme environmental adaptations. *PLoS ONE* 7:1–11. DOI: 10.1371/journal.pone.0034458.

Figure 1(on next page)

Evolutionary history of the genus *Exiguobacterium*

Evolutionary history of the genus *Exiguobacterium .* A) Phylogenetic reconstruction using complete genomic sequences of 17 representative *Exiguobacterium* strains. The tree was built with PhyloPhlAn (Segata et al., 2013) . B) **Synteny among *Exiguobacterium* strains**. Nucleotide syntenic blocks are represented by colored bars. Red links denote no rearrangements between the blocks compared. Blue links denote rearrangements between the blocks compared. Blue numbers in the phylogeny denote the minimum number of rearrangements obtained with *MGR*. Plasmids from *E. sibiricum* are displayed at the right (separated from the chromosome by backslashes). Black numbers indicate bootstrap values different from 100%.

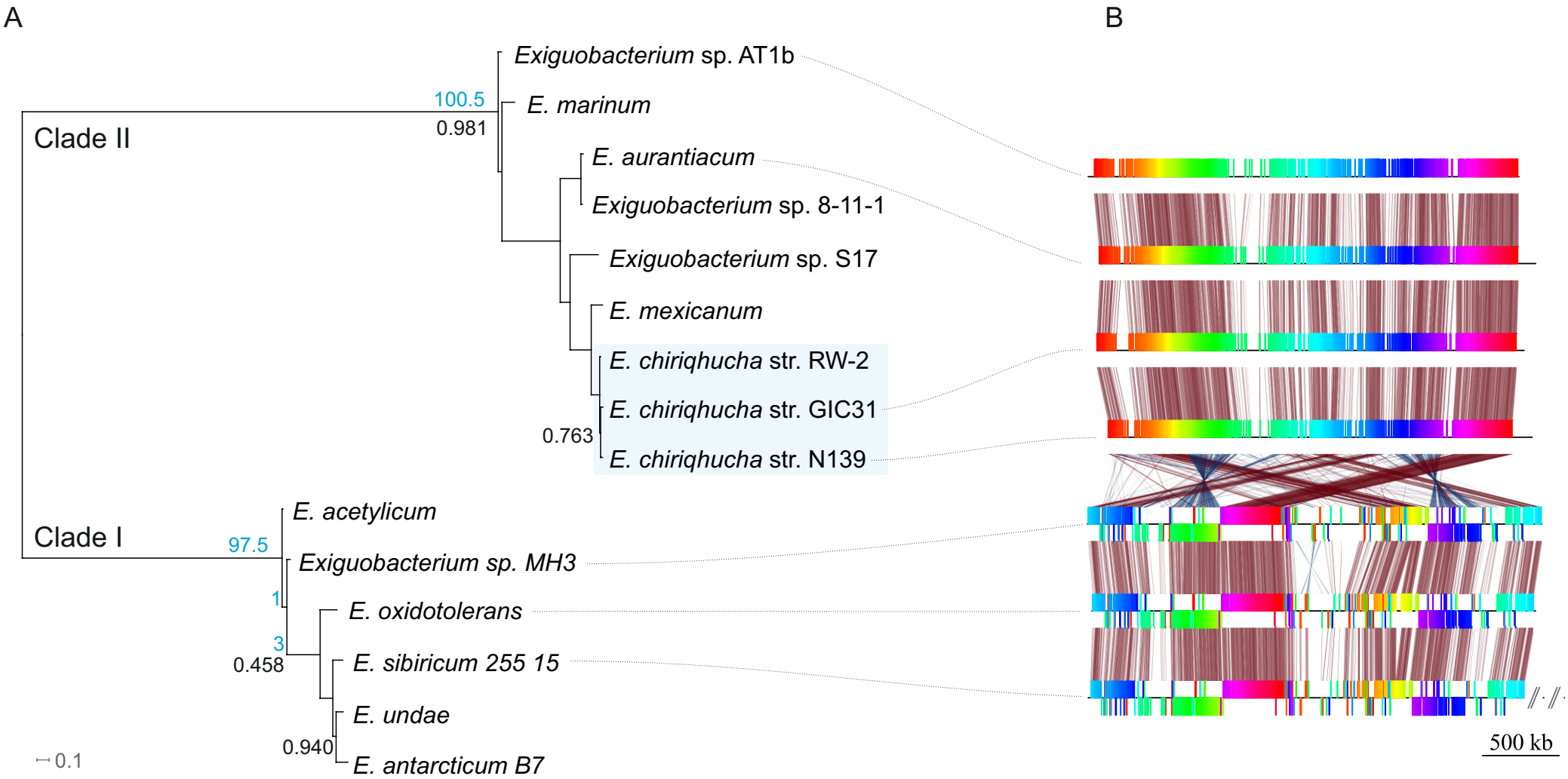
A
B
Exiguobacterium sp. AT1b
E. marinum
100.5
0.981
Clade II
E. aurantiacum
Exiguobacterium sp. 8-11-1
Exiguobacterium sp. S17
E. mexicanum
E. chiriqhucha str. RW-2
E. chiriqhucha str. GIC31
0.763
E. chiriqhucha str. N139
E. acetylicum
Clade I
97.5
Exiguobacterium sp. MH3
1
E. oxidotolerans
3
0.458
E. sibiricum 255 15
E. undae
0.940
E. antarcticum B7
0.1
500 kb

Figure 2

Differential interference contrast image of *E. pavilionensis* str. N139

**Note: Auto Gamma Correction was used for the image. This only affects the reviewing manuscript. See original source image if needed for review.*

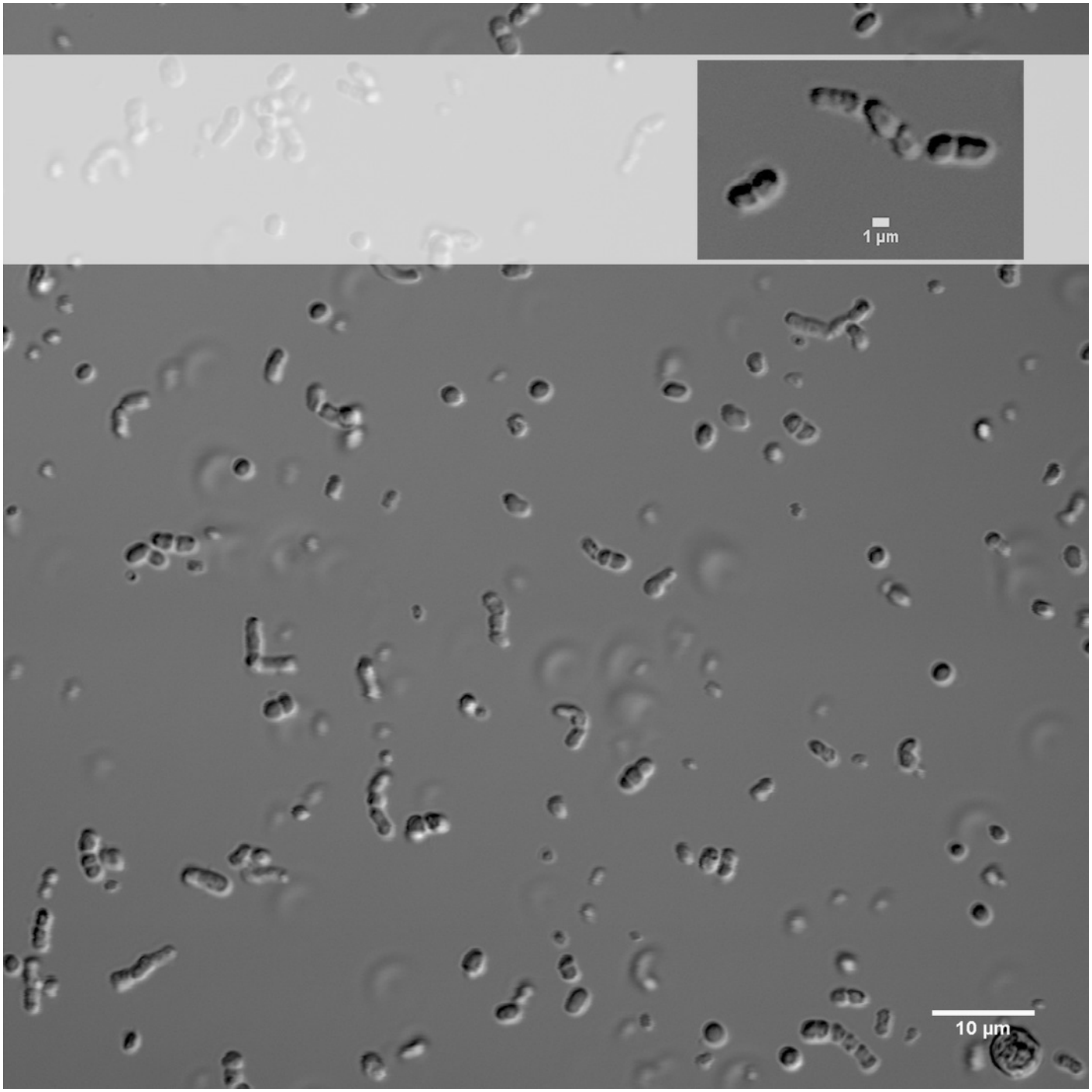

Figure 3(on next page)

Circular genome map of *E. pavilionensis* str. N139

Circle tracks from out towards inside are as follows: 1) Length in nucleotides for each contig; 2) Coding Sequences (CDS) in the Forward Strand (light blue); 3) CDS in the reverse strand (dark blue); 4) Strain Specific Genes (SSGs) in the forward strand (light purple); 5) SSGs in the reverse strand (dark purple); 6) GC Skew (gray). Skew and gene distribution follow that of a typical Firmicute genome. The Strain Specific Genes in the contigs that belong to the chromosome appear to be randomly distributed, whilst they seem to be concentrated in the contigs 12 and 13, which are probably the ones belonging to megaplasmids. The circular plot was done with Circos software (Krzywinski et al., 2009).

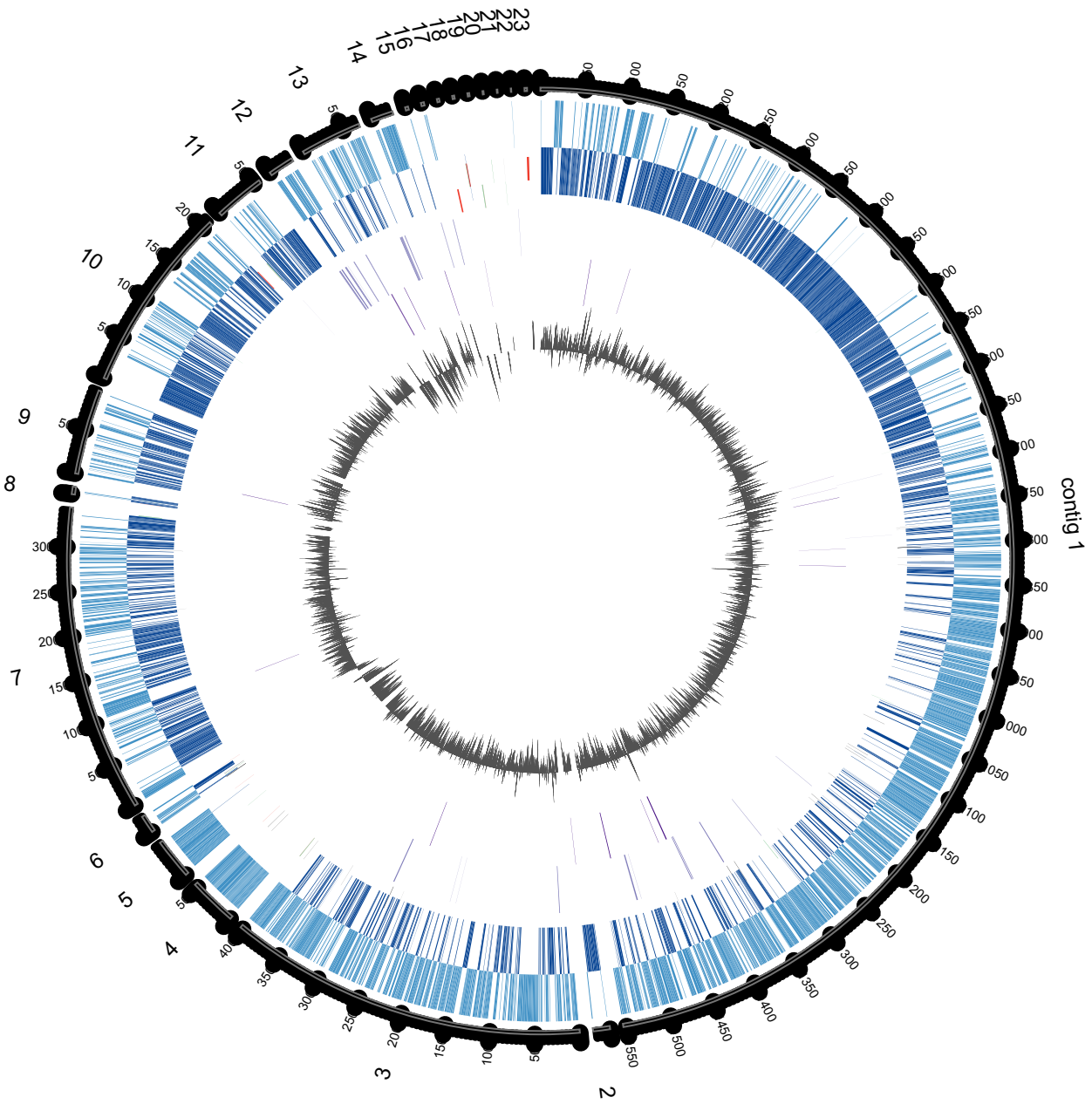
contig 1

Figure 4

Effect of ultraviolet B (UV-B) radiation on *Exiguobacterium* strains

Percentage survival to UV-B radiation of str. N139 (dark circle), str. S17 (light circle) and str. DSMZ 6208 (dark triangle). The influence of UV-B radiation was studied by exposing liquid cultures to increasing doses, varying exposure times between 0 and 240 minutes.

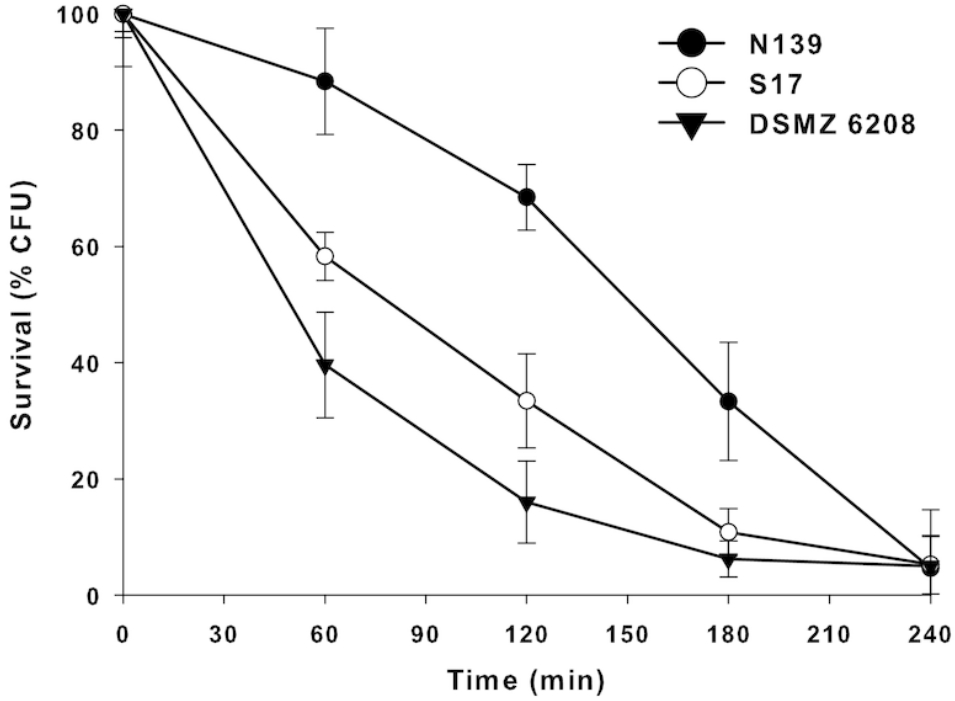
N139
S17
DSMZ 6208
Survival (% CFU)
100
80
60
40
20
0
0
30
60
90
120
150
180
210
240
Time (min)

Table 1(on next page)

Classification and general features of *Exiguobacterium pavilionensis* str. N139

[a] Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence); EXP: Inferred from Experiment. These evidence codes are from the Gene Ontology project ("Gene Ontology Evidence Codes")

Table 1. Classification and general features of *Exiguobacterium chiriqhucha* str. N139

Property	Term	Evidence code[a]
Classification	Domain *Bacteria*	TAS (Woese, Kandler & Wheelis, 1990)
	Phylum Firmicutes	TAS (Gibbons & Murray, 1978)
	Class Bacilli	TAS (De Vos et al., 2009)
	Order Bacillales	TAS (De Vos et al., 2009)
	Family Bacillales *Family XII. Incertae Sedis*	TAS (De Vos et al., 2009)
	Genus Exiguobacterium	TAS (De Vos et al., 2009; Vishnivetskaya, Kathariou & Tiedje, 2009)
	Species *Exiguobacterium chiriqhucha*	TAS (White III, Grassa & Suttle, 2013)
	Strain: *N139 (Accession: JMEH00000000)*	
Gram stain	*Positive*	IDA
Cell shape	*Short rods*	IDA
Motility	*Motile*	IDA
Sporulation	*Non-sporulating*	EXP
Temperature range	*Mesophilic (30 - 37°C)*	IDA
Optimum temperature	*30°C*	IDA
pH range; Optimum	*7-9*	IDA
Carbon source	*β-Methylglucoside, Galacturonic acid, L-asparagine, Tween 40, L-Serine, N-acetyl-glucosamine, Hydroxybutyric acid, Itaconic acid, Ketobutyric acid, Putrescine (See table S1)*	EXP
Habitat	*Aquatic*	TAS (Flores et al., 2009)
Salinity	*0.11% - 10% NaCl (w/v)*	IDA
Oxygen requirement	*Facultatively anaerobic*	TAS (De Vos et al., 2009)
Biotic relationship	*free-living*	IDA
Pathogenicity	*non-pathogen*	NAS
Geographic location	*Laguna Negra, Catamarca, Argentina*	IDA
Sample collection	*2006*	IDA
Latitude	*27°39'20.17"S*	IDA
Longitude	*68°33'46.18"W*	IDA
Altitude	*4100 masl*	IDA

[a] Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence); EXP: Inferred from Experiment. These evidence codes are from the Gene Ontology project ("Gene Ontology Evidence Codes")

Table 2(on next page)

Project information

1 **Table 2.** Project information.

Property	Term
Finishing quality	Permanent-draft
Libraries used	454 pyrosequence standard library
Sequencing platforms	454 Titanium
Fold coverage	85 x
Assemblers	Newbler 2.8 and MIRA 3.4
Gene calling method	Prokka
Locus Tag	EF88
Genbank ID	JMEH00000000.1
GenBank Date of Release	December, 2015
GOLD ID	Go0093977
BIOPROJECT	PRJNA245187
Source Material Identifier	N139
Project Relevance	UV resistance, metal resistance, adaptation to oligotrophic environments

2

Table 3(on next page)

Nucleotide content and gene count levels of the *E. pavilionensis* str. N139 genome

a) The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.

Table 3. Nucleotide content and gene count levels of the *E. chiriqhucha* str. N139 genome

Attribute	**Genome (total)**	
	Value	% of total[a]
Genome size (bp)	2,952,588	-
DNA coding (bp)	2,655,834	89.94
DNA G+C (bp)		52
DNA Scaffolds	23	
Total genes	3,182	100
RNA genes	82	2.62
Protein-coding genes	3,049	95.82
Pseudogenes	26	0.81
Genes in internal clusters	NA	
Genes with function prediction	2,356	74.04
Genes assigned to COGs	2,575	80.92
Genes with Pfam domains	2,538	79.76
Genes with signal peptides	NA	
Genes with transmembran helices	888	27.90
CRISPR repeats	0	

a) The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.

Table 4(on next page)

Genes associated with the 25 general COG functional categories

a) The total is based on the total number of protein coding genes in the annotated genome.

Table 4. Genes associated with the 25 general COG functional categories

Code	Value	% of total[a]	Description
J	166	5.44	Translation, ribosomal structure and biogenesis
A	0	0	RNA processing and modification
K	235	7.71	Transcription
L	144	4.72	Replication, recombination and repair
B	1	0.03	Chromatin structure and dynamics
D	36	1.18	Cell cycle control, Cell division, chromosome partitioning
V	62	2.03	Defense mechanisms
T	166	5.44	Signal transduction mechanisms
M	144	4.72	Cell wall/membrane biogenesis
N	75	2.46	Cell motility
U	53	1.74	Intracellular trafficking and secretion
O	100	3.28	Posttranslational modification, protein turnover, chaperones
C	152	4.99	Energy production and conversion
G	232	7.61	Carbohydrate transport and metabolism
E	224	7.35	Amino acid transport and metabolism
F	84	2.76	Nucleotide transport and metabolism
H	97	3.18	Coenzyme transport and metabolism
I	81	2.66	Lipid transport and metabolism
P	170	5.58	Inorganic ion transport and metabolism
Q	54	1.77	Secondary metabolites biosynthesis, transport and catabolism
R	463	15.19	General function prediction only
S	327	10.72	Function unknown
-	447	15.55	Not in COG

[a]The total is based on the total number of protein coding genes in the annotated genome.