Exiguobacterium pavilionensis str. N139, a halotolerant, UV-B and metal resistant bacteria from a high-altitude Andean lake

Ana Gutiérrez-Preciado[1†], Carlos Vargas-Chávez[1†], Mariana Reyes-Prieto[1], Omar F Ordoñez[2], Diego Santos-Garcia[1], 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, Apdo, Postal 510-3, Cuernavaca, Morelos 62250, México

[6]Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apdo, Postal 510-3, Cuernavaca, Morelos 62250, México

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 pavilionensis* str. N139, isolated from a high-altitude Andean lake. The 2,952,588-bp genome contains one chromosome and three megaplasmids. The genome analysis suggests the presence of enzymes that confer *E. pavilionensis* str. N139 the ability to grow under multiple environmental extreme conditions, including high concentrations of different metals and high ultraviolet B radiation. 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. pavilionensis* str. N139

Keywords

Exiguobacterium, High Altitude Andean Lakes, UV resistance, Arsenic resistance, bacterial metabolism, tryptophan biosynthesis, Extremophiles, Metals or Metalloids

Abbreviations

HAALs: high altitude Andean Lakes   ANI: Average Nucleotide Identity   LM: Lake Medium

SCCore: Single Copy Core   SSGs: Strain Specific Genes   HGT: Horizontal Gene Transfer

BCAA: Branched Chain Amino Acids   PER: Photoenzymatic Repair

NER: Nucleotide Excision Repair   PRR: Post Replication Repair

Introduction

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).

73   The *Exiguobacterium pavilionensis* str. N139 was selected for genome sequencing due to

74   its stress defense mechanisms such as its tolerance to high UV-B radiation, salinity and

75   metalloids, particularly arsenic. This strain was isolated from the water column of Laguna Negra,

76   which belongs to the 'Salar de la Laguna Verde', a system of five shallow oligotrophic lakes

77   originated in the Tertiary (65 million to 1.8 million years ago).

78   In the present study we characterized the genome of *E. pavilionensis* str. N139, in order

79   to identify the strategies that this organism employs to cope with the extreme environmental

80   factors present in the aforementioned lake, mainly those related to metal and UV-B resistance.

Classification and features

82   Members of the Exiguobacterium genus are Firmicutes, Gram-positive, facultative

83   anaerobes with a low G + C content (Vishnivetskaya, Kathariou & Tiedje, 2009).

84   *Exiguobacterium* is widely distributed all over the world (Karami et al., 2011) and has been

85   isolated and typified from a wide variety of environments including hot springs (Vishnivetskaya,

86   Kathariou & Tiedje, 2009; Vishnivetskaya et al., 2011), hydrothermal vents (Crapart et al.,

87   2007), permafrost (Vishnivetskaya & Kathariou, 2005; Vishnivetskaya et al., 2006; Rodrigues et

88   al., 2008), marine sediment (Kim et al., 2005), oligotrophic environments (Rebollar et al., 2012),

89   biofilms (Carneiro et al., 2012), alkaline methanogenic microcosms (Rout, Rai & Humphreys,

90   2015) and more recently in water and microbial mats from high-altitude desert wetlands

91   (Ordoñez et al., 2013). The Exiguobacterium genus is divided in two main phylogenetic clades

92   (Vishnivetskaya, Kathariou & Tiedje, 2009); clade I is composed of temperate and cold-adapted

93   strains, whereas clade II includes alkaliphilic species, with a marine origin and/or from high-

94   temperature habitats (Figure 1A).

95   *E. pavilionensis* str. N139, which belongs to clade II, was isolated from the water column

96   of Laguna Negra, in the HAALs (GPS: 27°38'49" S, 68°32'43" W) and in laboratory conditions

97   can uptake a wide variety of carbon sources (Table S1). Its cells are short rods and do not

98   sporulate (Figure 2, Table 1).

Materials and Methods

99

Genome project history

100

101   Our group is currently sequencing several *Exiguobacterium* strains isolated from different

102   oligotrophic ecosystems including different lakes in the Cuatrociénegas basin, México and in the

103   HAALs, Argentina, in order to understand how these bacteria adapt to poor nutrient availability.

104   In addition, the *E. pavilionensis* str. N139 was selected for its observed ability to cope with high

105   UV-B radiation and metal resistance (data not shown).

Growth conditions and genomic DNA preparation

106

107   *E. pavilionensis* str. N139 was isolated from Laguna Negra by plating it in Lake Medium

108   (LM). LM was used to maintain the same salinity as the isolation environment and was obtained

109   by filtering lake water (0.22 µm Biopore filters) and adding 2.5 g of yeast extract and 12 g of

110   agar (Difco) per liter at 20 °C. For future assays the strain was grown in LM broth at 20 °C with

111   agitation. DNA was extracted using the protocol described by (Fernández-Zenoff, Siñeriz &

112   Farías, 2006).

Microscopy

113

114   Differential interference contrast (DIC) images were obtained from cells grown on LB

115   medium overnight, and mounted in No. 2 coverslips (Figure 2). LB medium was used as

116 mounting media during image acquisition. Images were shot with an Olympus FV1000 Laser

117 Scanning Confocal on an Olympus IX81 inverted microscope equipped with 60x UPlanSApo

118 NA 1.3 Sil objective lens. With a 405nm laser line, DIC Images were acquired in the TD channel

119 controlled with Olympus FV10-ASW-4.2 software. Brightness, contrast and scale bars were

120 adjusted on displayed images using the Fiji software.

Phylogenetic Reconstruction

122 For the phylogenetic analysis 16S rRNA gene sequences for all *Exiguobacterium* strains

123 whose genome sequences have been completed were used (*Exiguobacterium acetylicum, E.*

124 *antarcticum, E. marinum, E. oxidotolerans, E. sibiricum* 255-15, *Exiguobacterium* sp. 8-11-1,

125 *Exiguobacterium* sp. AT1b, *Exiguobacterium* sp. GIC31, *E. pavilionensis* str. RW-2,

126 *Exiguobacterium* sp. S17 and *E. undae)*. Furthermore, the 16S sequences from other

127 *Exiguobacterium* strains relevant for this analysis (*E. arabatum, E. mexicanum*), and from other

128 Firmicutes that were used as outgroups (*Bacillus anthracis* str. AMES, *Oceanobacillus iheyensis,*

129 *Bacillus subtilis* str. 168 and *Bacillus hablodurans* str. C125) were also included. All 16S rRNA

130 sequences were downloaded from the SILVA database, with the exception of the 16S rRNA

131 sequence from *Exiguobacterium* sp. MH3 which was downloaded from NCBI. All sequences

132 were manually checked for length and quality; only high quality sequences, longer than 1,400 bp

133 were included in the analysis. Sequences were aligned with *ssu-align* and then masked with *ssu-*

134 *mask* (Nawrocki, Kolbe & Eddy, 2009). Positions of the bases that are part of stems or loops

135 according to their secondary structure were calculated with *esl-alistat* from the *Infernal* package

136 to obtain a better alignment (Nawrocki & Eddy, 2013). The dataset was partitioned into stems

137 and loops and treated under different evolutionary models. *Ad hoc* scripts were written to

138 integrate all the position's information in a final nexus file alignment. The resulting file was used

139   as the input for the Bayesian phylogenetic reconstruction with *MrBayes* v 3.2.5 (Ronquist &

140   Huelsenbeck, 2003). Substitutions in stem regions were modeled considering the co-variation of

141   the paired nucleotides, with the doublet model (Schöniger & Von Haeseler, 1994). For the loop

142   regions of the 16S rRNA, a traditional 4×4 model was used. With the data partitioned in this

143   fashion, and the whole gene being subjected to a GTR model with invariable gamma, two sets of

144   3,000,000 iterations were run, sampling and saving every 100th tree to a file, with the first 250

145   trees discarded. A majority-rule consensus was used to examine estimates of posterior

146   probabilities, interpreted as nodal support. Figure 1A shows the phylogenetic reconstruction of

147   all *Exiguobacterium* species from both clades. These analyses placed the strain N139 as most

148   similar to *E. pavilionensis* str. RW-2 (White III, Grassa & Suttle, 2013), which is coherent with

149   the average nucleotide identity (ANI) analyses (see below). ANI calculations were done for

150   N139 *versus* all other complete genomic sequences of *Exiguobacterium*, and these two strains,

151   along with *Exiguobacterium* sp. GIC31, belong to the same species (Table S2).

152   **Genome sequencing and assembly**

153   The genome of *E. pavilionensis* str. N139 was generated using *454 technology* (Table 2).

154   A standard *454 Titanium* library was constructed and sequenced, producing 664,086 reads,

155   totaling 253.9 Mb of data. The 454 data was assembled with *Newbler*, version 2.8 and MIRA,

156   version 3.4 (Chevreux et al., 2004). Both assemblies were merged using *Minimus2*, from the

157   *amos* version 3.1.0, with assembly errors manually corrected. The contigs were sorted with

158   *Mauve* version 2.3.1 (Rissman et al., 2009)*,* using *Exiguobacterium* sp. AT1b as the reference

159   because it is the closest relative with a completely sequenced genome (Vishnivetskaya et al.,

160   2011). The final assembly consists of 23 contigs, the smallest one being 457 bases in length and

161   the largest 1.5 Mb, with an average coverage of 85×.

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 COG, a multiple sequence alignment of *bona fide* representative sequences were generated using the *Muscle* program (Edgar, 2004), and then, the corresponding Hidden Markov Model was built using the *hmmerbuild* program, also provided in the *HMMER3* package (Mistry et al., 2013). The resulting annotation was subjected to manual curation. *Pathway Tools 13* (Karp et al., 2009) in combination with the *BioCyc* (Caspi et al., 2014) and *UNIPROT* (Consortium, 2015) databases were used to infer the metabolic capacities of *E. pavilionensis* str. N139. The curated model of *E. pavilionensis* str. N139 can be provided upon request, and will be deposited in the *BioCyc* database.

Results and Discussion

Genome Properties

The genome of *E. pavilionensis* str. N139 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.*

184 *pavilionensis* str. N139 has 10 ribosomal rRNA operons, confirmed by PFGE (see Supplemental

185 Material and Figure S2). A putative function was assigned to 2,214 (73%) of the protein-coding

186 genes, and the remaining genes were annotated as hypothetical proteins. The properties and the

187 statistics of the genome are summarized in Table 3. 2,575 protein-coding genes were assigned to

188 1,603 COG families, corresponding to a gene content redundancy of 38.1% (see Table 4).

189 **Genome Rearrangements**

190   To explore the genomic rearrangements present on *E. pavilionensis* str. N139 in

191 comparison to other *Exiguobacterium* species, nucleotide syntenic blocks were obtained with

192 *Mauve* version 2.3.1 (Darling et al., 2004). Syntenic block permutations were exported and used

193 as input for *MGR* (Bourque & Pevzner, 2002). *MGR* was used to calculate the minimum number

194 of rearrangements between the species analyzed, and to recover the rearrangement dendrogram.

195 *genoPlotR* (Guy, Kultima & Andersson, 2010) was used to plot the syntenic blocks (Figure 1B).

196 Genome rearrangements within clades I and II are scarce, showing high conservation of the

197 genomic structure within clades. However, several genomic rearrangements occurred as both

198 clades diverged.

199   In order to determine which contigs of *E. pavilionensis* str. N139 belong to plasmids, the

200 plasmid sequences of pEspA and pEspB from *E. arabatum* RFL1109 (Jakubauskas et al., 2009)

201 were retrieved from NCBI. This strain was selected for comparison because their plasmids have

202 been widely studied (Jakubauskas et al., 2009) and because it is phylogenetically close to *E.*

203 *pavilionensis* str. N139. Jakubauskas and colleagues identified the regions hr-A1, hr-AB and hr-

204 A2 in plasmid pEspA as capable of replicating the plasmid in *Bacillus*. For plasmid pEspB, they

205 hypothesized that the regions hr-B1, hr-AB and hr-B2 are involved in a theta replication

206 mechanism (Jakubauskas et al., 2009). BLAST searches of these regions were performed against

207 all *Exiguobacterium* genome sequences available to date. For the strains *E.* MH3, *E. antarcticum*

208 and *E.* AT1b, which are described as genomes without plasmids, no significant hits were found.

209 Conversely, hits to the *E. arabatum* sequences hr-B1, hr-AB-B and hr-B2 (a fragment of 39 kb)

210 were found in the genomes of *E.* GIC3 (56 kb) and *E.* N139 (contig000014 of size 25 kb). It was

211 concluded that the sequences present in the plasmids are shared within different

212 *Exiguobacterium* strains, displaying a highly dynamic behavior, and as a result, determining

213 which of our contigs correspond to the three megaplasmids observed in the PFGE experiments

214 was not possible (see Supplemental Material). Furthermore, contig 14 in our assembly

215 corresponds to the smallest contig of *E.* GIC3, so it could be a plasmid in both *Exiguobacterium*

216 strains. However, contig 14 lies adjacent to contig 13, both accounting for a total size of 100 kb

217 when synteny was evaluated against *E. pavillionensis* RW-2. It is worth mentioning that contig

218 13 possesses most of the genes responsible for metals resistance, but this region appears to be

219 integrated in the chromosome of *E.* GIC3. This highly dynamic behavior across strains suggests

220 that, if both contigs belong to a plasmid, it might be an integrative one. A MAUVE analysis

221 performed between the strains N139, GIC3 and *E. pavillionensis* RW-2 shows high synteny

222 across their chromosomes. This idea that contigs 12, 13 and 14 might belong to the plasmids is

223 supported by their shifts in GC skew (Figure 3). To all appearances, the chromosomes within

224 each of the two main clades of the *Exiguobacterium* species are very similar, but quite distinct

225 when compared between these clades (Figure 1).

226 **The *Exiguobacterium* strain N139 belongs to the *pavilionensis* species**

227   ANI calculations of all clade II *Exiguobacterium* strains were performed and compared to

228 our N139 strain, suggesting that *E. pavilionensis* str. RW-2, *Exiguobacterium* sp. GIC31 and this

229   N139 strain belong to the same species since they share ANI values above 97% (Table S2)

230   (Goris et al., 2007). Typically, the ANI values between genomes of the same species are above

231   95% (e.g., *E. coli*). Also relying on the ANI calculations, it was concluded that the outgroup of

232   the *E. pavilionensis* species could be *E. mexicanum*. To better assess if *E. mexicanum* belongs to

233   the same species, additional experimental support is required, since its genome is 400 genes

234   bigger than the remaining members of the *E. pavilionensis* clade.

235   **Exiguobacterium clade II pangenomes**

236   To further understand the genomic properties of the three *E. pavilionensis* strains, their

237   pangenome was built along with their putative outgroup, *E. mexicanum*. This pangenome

238   comprises all their 4,100 genes; 2,524 of them belonging to a Single Copy Core (SCCore)

239   Genome (Figure S3). Orthologs were first calculated following the OrthoMCL pipeline (Li, Jr &

240   Roos, 2003; Fischer et al., 2011), and the pangenome and the SCCore were elucidated using *ad*

241   *hoc* perl scripts. The same process was repeated for all *Exiguobacterium* strains from clade II

242   whose complete genomes were available at the time of analysis. The pangenome of the nine

243   *Exiguobacterium* from clade II is composed of 5,267 genes; 2,116 belonging to the SCCore and

244   1,664 being Strain Specific Genes (SSGs). The resulting pangenome shows a very conserved and

245   cohesive pool of genes, despite their evolutionary distance and their remote geographic locations.

246   Over two thousand genes compose the SCCore Genome, which represents a large core genome

247   when compared to other pangenomes (Lefebure & Stanhope, 2007; Rasko et al., 2008; Rouli et

248   al., 2015; Zhang & Qiu, 2015), and taking into account that the average genome size of

249   *Exiguobacterium* strains is approximately three thousand genes. SSGs could have been acquired

250   by Horizontal Gene Transfer (HGT) and retained to adapt to these diverse environments, or

251   equally likely, lost in some of the living taxa, due to lack of selective pressure in their respective

252   niches. The SSGs are represented in a heatmap on Figure S4 where the clusterization of the

253   *Exiguobacterium* strains is based on the presence (and abundance) or absence of their COG

254   assignation. *Exiguobacterium* sp. S17 and *E. mexicanum* are exceptional for the fact that they

255   possess a large pool of SSGs (Table S3).

256   Fifty-nine of the SSGs found in the *E. pavilionensis* str. N139 were mapped on its genome to

257   see if their distribution followed some bias (Figure 3). Throughout the contigs that are putative

258   chromosomal regions, the SSGs appear to be randomly distributed. However, some of the SSGs

259   are concentrated in the contigs 12 and 13, supporting the previous idea that these contigs may be

260   part of the megaplasmids seen in the PFGE analysis (Figure S1). COGs were assigned to the

261   SSGs as previously described for the *E. pavilionensis* str. N139 genome. For the whole set of

262   SSGs of the pangenome, COGs were successfully assigned to 66% of the genes, and are

263   represented in a heatmap (Figure S4). However, most of the N139 SSGs could not be assigned to

264   COGs, and for those that were successfully classified, the vast majority falls in the S and R

265   (Poorly Characterized) COG categories, leaving open questions on which may be the unique

266   strategies that N139 employs to adapt to the particular environment of Laguna Negra.

267   **Main Metabolic Pathways, Amino acids, Nucleotides and Cofactors**

268   Based on its genomic content, *E. pavilionensis* str. N139 is probably a chemoheterotroph

269   since it has two copies of *aioB* arsenite oxidase, which means it could obtain energy from

270   arsenite oxidation. It has the complete pathway for glycolysis and it could synthesize acetyl-

271   CoA, succinyl-CoA and isobutanoyl-CoA. It is a heterolactic fermenter, being able to produce

272   lactate from pyruvate and ethanol from acetaldehyde. It has a complete TCA cycle, and it lacks

273   the first two steps of the pentose phosphate pathway, but the rest of the pathway is present.

274   Hence, its central metabolism is similar to *B. subtilis* [48–50], but *E. pavilionensis* str. N139 can

275   synthesize more fermentation products, namely ethanol and formate. *E. pavilionensis* str. N139

276   lacks the routes for synthesizing *de novo* phenylalanine and tyrosine, as well as the Branched

277   Chain Amino Acids (BCAA). However, it can synthesize tyrosine from phenylalanine, since it

278   has the phenylalanine-4-hydroxylase regulated by a Tyr (UAC codon) T box riboswitch. Despite

279   lacking the complete pathways for BCAA biosynthesis, it preserves the *ilvE* gene, a BCAA

280   aminotransferase, which could probably synthesize any of the three BCAAs from available

281   precursors. An interesting note on its tryptophan biosynthesis is that its biosynthetic operon is

282   split in two transcription units: *trpEG* and *trpDCFBA*, which are separated in the chromosome,

283   but co-regulated by a Trp T box riboswitch. Although this regulation is common in Firmicutes

284   (Gutierrez-Preciado et al., 2005; Gutiérrez-Preciado, Yanofsky & Merino, 2007), the genome

285   context of the *trp* operon is not, and it is interesting that this separation takes place at the

286   synthesis of anthranilate. Moreover, the *trpEG* genes are regulated by a single T box, whilst the

287   *trpDCFBA* operon is regulated by two T boxes in tandem. This could either mean that the

288   separation of the pathway is a recent event and the regulation is being settled in order to

289   coordinate both transcriptional units; or that this strain requires anthranilate (the product of

290   *trpEG*) for something else. Certainly, one possibility is that *E. pavilionensis* str. N139 exports

291   anthranilate for a synthrophy with a partner(s) and the subsequent steps of the tryptophan

292   biosynthetic pathway require a stricter regulation in order for the genes *trpDCFBA* to be

293   expressed. Since *E. pavilionensis* str. N139 lacks the biosynthetic pathways for five amino acids,

294   a likely scenario is that this bacterium is sharing metabolites with other partners in Laguna

295   Negra. This is supported by the observation that it is able to form part of a biofilm (data not

296   shown), and that in all of the amino acids tested it can only grow on serine and asparagine as a

297   sole carbon source (see Table S1). Based on the metabolite tracer from Pathway Tools, it can be

298   inferred that *E. pavilionensis* str. N139 could synthesize phenylalanine as well as valine from

299   serine or asparagine. In the same fashion, it cannot grow with phenylalanine as the sole carbon

300   source. Therefore, the configuration of the genes involved in amino acid metabolism might

301   represent a requirement of amino acid syntrophy that needs further exploration and testing. A

302   second possibility is that *E. pavilionensis* str. N139 is utilizing anthranilate for some novel

303   pathway. Anthranilate cannot be in excess with respect to tryptophan, since its excess could

304   decrease the availability of phosphoribosyl pyrophosphate (PRPP) for histidine synthesis (and

305   other reactions) (Merino, Jensen & Yanofsky, 2008). This novel pathway could be involved in

306   different functions that require either tryptophan or anthranilate as intermediates. Examples of

307   these functions are quorum sensing molecules in *Pseudomonas aeruginosa* (Farrow & Pesci,

308   2007), to plant hormones in *Azospirillum brasilense* (Ge, Xie & Chen, 2006), violacein in

309   *Chromobacterium violacein* (Antônio & Creczynski-Pasa, 2004) or antibiotics as in

310   *Streptomyces coelicolor* (Amir-Heidari, Thirlway & Micklefield, 2008).

311   The regulation of biosynthetic and transporter genes through riboswitches is common in

312   Firmicutes, specially the members of Bacilli class. It has also been observed that transport and

313   biosynthesis of the same metabolite tend to be part of a regulon mediated by *in cis* elements, like

314   riboswitches (Gutiérrez-Preciado et al., 2009). Methionine can be synthesized and imported

315   through several strategies. Several SAM riboswitch regulated operons coding for Met

316   transporters were identified in the genome of *E. pavilionensis* str. N139, as well as canonical *met*

317   biosynthetic genes. An interesting case is the methionine salvage pathway, whose genes are

318   encoded in two divergent operons, both regulated by divergent SAM riboswitches. Both operons

319   must be transcribed in order for the Yang cycle to be completed. In one operon, genes *mtnK* and

320   *mtnA* are transcribed along with three ribose transporters, *rbsB, rbsC* and *araG*. Lysine

321 biosynthesis (from aspartate via diaminopimelate) and transport are part of a regulon under the

322 lysine riboswitch. Furthermore, through the identification of riboswitches, two transporters from

323 the NhaC family were annotated: one as a methionine transporter (SAM riboswitch), and the

324 other one as a lysine transporter (LYS riboswitch). This strategy of improving gene annotation

325 through the knowledge of the gene's regulation has been previously explored (Rollins, 2002;

326 Gutiérrez-Preciado & Merino, 2012; Gutiérrez-Preciado et al., 2015).

327 **Cofactors.** Thiamine can be synthesized *de novo*, its biosynthesis and its uptake are

328 regulated by the TPP riboswitch. Moreover, the analysis of *E. pavilionensis* str. N139 genome

329 indicates that a new thiamine transporter could be present in this bacterium. The gene

330 exiN139_02072 is automatically identified as a membrane protein, but it seems to be regulated

331 by a TPP riboswitch. Experimental evidence is needed for the confirmation and characterization

332 of this transporter, which could unveil a new family of thiamine transporters. Riboflavin

333 biosynthesis and transport (RibU) are also co-regulated through a FMN riboswitch.

334 **Nucleotides.** In the genome of *E. pavilionensis* str. N139 the purine *de novo* biosynthetic

335 pathway is encoded in a huge transcription unit regulated by a purine riboswitch. Other

336 transcription units in the same regulon include a monocystronic GMP synthase, and genes

337 involved in adenine and adenosine salvage pathway.

338 ## Genomic Adaptations to an Extreme Environment
339 Laguna Negra is an aquatic ecosystem that harbors extreme environmental conditions

340 such as high levels of UV-B (10.65 wm2), high salinity levels (32%), scarce nutrients,

341 particularly phosphorous (<005 mg/l), high metal contents including the metalloid arsenic

342   (3mg/l), an alkaline pH and large daily temperature fluctuations (ranging from 20 °C during the

343   day to −40 °C at night) (Flores et al., 2009); (see Table 1).

Resistance to metals and metalloids

345   In Laguna Negra, ubiquitous Arsenic enters the *E. pavilionensis* str. N139 cells through

346   existing transporters due to its high structural similarity with other molecules (Rosen, 1999) and

347   induces oxidative stress responses (Oremland & Stolz, 2003). Furthermore, arsenite (AsO_2H) and

348   arsenate (AsO_4^{3-}), are both toxic molecules. Arsenite binds to reduced cysteines in proteins

349   inactivating them, and arsenate is a molecular analog of phosphate and therefore inhibits

350   oxidative phosphorylation (Oremland & Stolz, 2003). Arsenate is far less toxic than arsenite,

351   hence the oxidation of arsenite is considered a detoxification process. However, the oxidation of

352   arsenite to arsenate, when coupled to the reduction of oxygen to water, is an exergonic process,

353   and it has been suggested that at least some bacteria may derive energy out of this process

354   (vanden Hoven & Santini, 2004). *E. pavilionensis* str. N139 has an arsenite oxidase, *aioB*,

355   enabling it to oxidize arsenite. This is an important metabolic capability, because it uses arsenite

356   as an electron donor. Moreover, from a bioremediation point of view, this former metabolic

357   feature is important since arsenite is more soluble than arsenate, so it can facilitate the removal

358   of As in solution. *E. pavilionensis* str. N139 also has an arsenite efflux pump, *arsB*, as well as an

359   ATPase that provides energy to ArsB for extrusion of arsenite and antimonite, *arsA*, co-

360   transcribed with *arsD*, an arsenic chaperone for the ArsAB pump (Páez-Espino et al., 2009).

361   Hence, this bacterium can probably detoxify and extrude As, as well as oxidize arsenite

362   acquiring energy from this process. *E. pavilionensis* str. N139 also possesses redundancy for

363   mercury detoxification, harboring four paralogous copies of *merA*. Briefly, MerA is the key

364   detoxification enzyme of the mercury resistance system, reducing Hg^{2+} to Hg^0 (Silver & Phung,

2005). Hg is toxic due to its high affinity to sulfur (Nies, 2003) and usually, *mer* resistance genes are co-transcribed in an operon whose dissemination is common by horizontal gene transfer (HGT) (Barkay, Miller & Summers, 2003). In this organism, two copies of *merA* are present in a monocystronic fashion; a third one is transcribed with a hypothetical protein. A fourth copy is co-transcribed with *merR*, the regulatory protein of the system. Two *mer* transporters which uptake Hg and *merP*, a transporter with a Sec-type signal, which could import Hg as a neutral chloride or hydroxide and deliver it to the other Mer transporters, which will finally transfer it to MerA.

The most common mechanism of resistance to metals consists of efflux pumps for inorganic ions. However, As and Hg resistance mechanisms are unique in the sense that these elements are reduced to lower their toxicity [66], instead of being exported. *E. pavilionensis* str. N139 is resistant to cadmium, zinc, cobalt, and copper by pumping it out from the cell. It has two membrane embedded Cd^{2+} efflux pumps, one of which can also extrude zinc and cobalt; two paralogous copies of *copA* and *copB*, two P-type ATPase systems for exporting copper, and *cueR,* a sensing cytoplasmic Cu that protects periplasmic proteins from copper-induced toxicity. *copB* is transcribed monocystronically, and each of the *copA* genes form an operon co-transcribed with a copy of *copZ*, a copper chaperone, but one is co-transcribed with a glutaredoxin, whilst the other is co-transcribed with *csoR*, a copper-sensitive operon repressor.

Resistance to UV radiation

Another extreme environmental condition in Laguna Negra is high ultraviolet radiation, particularly UV-B (Flores et al., 2009; Ordoñez et al., 2009). Bacteria have different UV damage repair pathways, including photoenzymatic repair (PER), nucleotide excision repair (NER) also called dark repair, and recombinational repair (PRR) or post-replication repair (Goosen &

388   Moolenaar, 2008). *E. pavilionensis* str. N139 has three genes (exiN139_00335 (*phrB*),

389   exiN139_01768 and exiN139_00235) related to photolyases, which are involved in PER. They

390   use UV as energy source (using FADH and transferring electrons) and catalyze the

391   monomerization of cyclobutyl pyrimidine dimers. The gene exiN139_00335 only has

392   homologues in Firmicutes including other known *Exiguobacterium,* and exiN139_01768 has

393   homologues in Firmicutes, Cyanobacteria, α- and γ- Proteobacteria, and Euryarcheotes.

394   Additionally, exiN139_00235 is a cryptochrome, which are flavoproteins related to photolyases.

395   Cryptochromes do not repair DNA and are presumed to act in other (unknown) processes, such

396   as entraining circadian rhythms (Yuan et al., 2012).

397   *E. pavilionensis* str. N139 has also genes for NER. Its genome encodes the UvrABC

398   endonuclease, a complex that recognizes DNA damage, binds to the damaged segment and

399   cleaves it. Additionally, it codes for PcrA (also known as UvrD), a helicase in charge of

400   removing the excised segment recognized and cleaved by UvrABC. These genes are regulated by

401   the SOS response, which uses LexA as a repressor inactivated by RecA (Minko et al., 2001).

402   Regarding the PRR, *E. pavilionensis* str. N139 encodes for RecA, which recognizes SSB

403   and cleaves UmuD, which becomes UmuD' and binds UmuC to generate polymerase V, which in

404   turn repairs damages, sometimes causing mutations. The gene exiN139_03003 may produce

405   polymerase IV that is also involved in DNA damage repair (Sommer, Bailone & Universitaire,

406   1998). UmuC is found in the genome, however UmuD is missing. It is possible that a protein

407   highly similar to an existing copy of LexA may be taking its role, given that both can be cleaved

408   by RecA and are present in *Exiguobacterium*.

409    Finally, it is likely that *E. pavilionensis* str. N139 participates in biofilm formation in

410 Laguna Negra along with other bacteria (unpublished results). Analyses of other

411 *Exiguobacterium* have shown that they participate in marine biofilms interacting with other

412 Firmicutes and Proteobacteria (López et al., 2006; Carneiro et al., 2012). Evidence of possible

413 biofilms associated genes originates from two *loci* present in the genome of *E. pavilionensis* str.

414 N139. The first *locus* encoding a protein capable of producing alginate, a linear co-polymer of

415 two uronic acids that is produced in its acetylated form by some bacteria for adherence of these

416 bacteria to target cell walls by the creation of a biofilm (Ramphall & Pier, 1985). The second

417 *locus* codes for the arginine deiminase system, which can function at very low pH and is thought

418 to be a critical factor in oral biofilm pH homeostasis (Burne & Marquis, 2000).

Conclusions

419

420   *E. pavilionensis* str. N139 lives in a high-altitude, salted lake exposed to intense UV

421 radiation, about 300 km away from the nearest ocean, the Pacific. Many factors in *E.*

422 *pavilionensis* str. N139 metabolism, such as the fact that it needs to uptake certain intermediates

423 like phenylalanine and BCAAs, and the possible excretion of the overproduced anthranilate,

424 suggest that it is a key player in the amino acid metabolism of a microbial consortium that

425 inhabit Laguna Negra. Moreover, the excess of anthranilate that it may produce could be directed

426 to some novel pathway that remains to be uncovered, such as a new antibiotic, a new pigment or

427 a new quorum sensing molecule.

428    Additionally, this microorganism lives in a low phosphorous environment, and relies on

429 strategies for phosphorous uptake, like the presence of high-affinity Pi transporters and its

430 regulation (*pstS, pstCAB, phoB, phoR, dedA* and *ptrA*) and genes for polyphosphate storage and

431 breakdown (*ppk* and *ppx*). Organisms that scavenge phosphate can sometimes uptake the

432   structurally similar arsenate ion, and hence also depend on arsenate detoxification mechanisms.

433   This *Exiguobacterium* is not only able to detoxify this metalloid but also metals like mercury,

434   cadmium, zinc, cobalt, and copper; and it has a complete defense system against UV damage. It

435   is also able to thrive in the alkaline environment of Laguna Negra since its genome code for all

436   the typical antiporters present in alkaliphilic bacilli (*nhaC, nhaP,* the *mpr* operon*, yhaU, norM*

437   and *mleN*). These antiporters present also contribute to a moderate salinity resistance (Ventosa,

438   Nieto & Oren, 1998; Padan et al., 2005), and this could also be related with the maintenance of

439   metal resistance strategies in its genome, since it has been shown that lowering the salinity can

440   lead to enhanced sensitivity to cadmium, cobalt and copper (Ventosa, Nieto & Oren, 1998). With

441   all these characteristics, *E. pavilionensis* str. N139 is an excellent candidate for future

442   biotechnological research.

443   Although our study generates more questions than the ones it could solve, by sequencing

444   its genome we have gained insights on the strategies the strain N139 employs for thriving in its

445   habitat. From its Strain Specific set of genes, only 23 out of 59 could be annotated and classified

446   to a COG, and still, most of the COG-classified genes belong to the poorly characterized

447   category. This set of genes of unknown function require further experimental work to completely

448   unveil how the strain N139 is adapting to the extreme environment of Laguna Negra.

449

450   Acknowledgements
451   We thank Jerome Verleyen for computer support and the Instituto de Biotecnología-

452   UNAM for allowing us access to its computer cluster. We would also like to thank Eric

453   Nawrocki for his guidance with the *ssu-align* program.

Competing interests

The authors declare that they have no competing interests.

References

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

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

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

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

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

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

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

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

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

Carneiro AR., Ramos RTJ., Dall'Agnol H., Pinto AC., Soares SDC., Santos AR., Guimarães LC., Almeida SS., Baraúna RA., das Graças DA., Franco LC., Ali A., Hassan SS., Nunes

489   CIP., Barbosa MS., Fiaux KK., Aburjaile FF., Barbosa EGV., Bakhtiar SM., Vilela D.,
490   Nóbrega F., dos Santos AL., Carepo MSP., Azevedo V., Schneider MPC., Pellizari VH.,
491   Silva A., de Castro Soares S., Santos AR., Guimarães LC., Almeida SS., Baraúna RA., das
492   Graças DA., Franco LC., Ali A., Hassan SS., Nunes CIP., Barbosa MS., Fiaux KK.,
493   Aburjaile FF., Barbosa EGV., Bakhtiar SM., Vilela D., Nóbrega F., dos Santos AL., Carepo
494   MSP., Azevedo V., Schneider MPC., Pellizari VH., Silva A. 2012. Genome sequence of
495   Exiguobacterium antarcticum B7, isolated from a biofilm in Ginger Lake, King George
496   Island, Antarctica. *Journal of Bacteriology* 194:6689–90. DOI: 10.1128/JB.01791-12.

497   Caspi R., Altman T., Billington R., Dreher K., Foerster H., Fulcher C a., Holland T a., Keseler
498   IM., Kothari A., Kubo A., Krummenacker M., Latendresse M., Mueller L a., Ong Q., Paley
499   S., Subhraveti P., Weaver DS., Weerasinghe D., Zhang P., Karp PD. 2014. The MetaCyc
500   database of metabolic pathways and enzymes and the BioCyc collection of
501   Pathway/Genome Databases. *Nucleic Acids Research* 42:459–471. DOI:
502   10.1093/nar/gkt1103.

503   Chevreux B., Pfisterer T., Drescher B., Driesel AJ., Müller WEG., Wetter T., Suhai S. 2004.
504   Using the miraEST assembler for reliable and automated mRNA transcript assembly and
505   SNP detection in sequenced ESTs. *Genome Research* 14:1147–1159. DOI:
506   10.1101/gr.1917404.

507   Consortium TU. 2015. UniProt : a hub for protein information. 43:204–212. DOI:
508   10.1093/nar/gku989.

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

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

515   Dib J., Motok J., Fernández-Zenoff V., Ordoñez O., Farías ME. 2008. Occurrence of resistance
516   to antibiotics, UV-B, and arsenic in bacteria isolated from extreme environments in high-
517   altitude (above 4400 m) Andean wetlands. *Current microbiology* 56:510–7. DOI:
518   10.1007/s00284-008-9103-2.

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

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

524   Farrow JM., Pesci EC. 2007. Two distinct pathways supply anthranilate as a precursor of the
525   Pseudomonas quinolone signal. *Journal of Bacteriology* 189:3425–3433. DOI:
526   10.1128/JB.00209-07.

527   Fernández-Zenoff V., Heredia J., Ferrero M., Siñeriz F., Farías ME. 2006. Diverse UV-B

528   resistance of culturable bacterial community from high-altitude wetland water. *Current*
529   *microbiology* 52:359–62. DOI: 10.1007/s00284-005-0241-5.

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

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

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

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

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

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

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

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

553   Gutiérrez-Preciado A., Henkin TM., Grundy FJ., Yanofsky C., Merino E. 2009. Biochemical
554   features and functional implications of the RNA-based T-box regulatory mechanism.
555   *Microbiology and molecular biology reviews MMBR* 73:36–61. DOI:
556   10.1128/MMBR.00026-08.

557   Gutierrez-Preciado A., Jensen R a., Yanofsky C., Merino E. 2005. New insights into regulation
558   of the tryptophan biosynthetic operon in Gram-positive bacteria. *Trends in Genetics*
559   21:432–6. DOI: 10.1016/j.tig.2005.06.001.

560   Gutiérrez-Preciado A., Merino E. 2012. Elucidating metabolic pathways and digging for genes of
561   unknown function in microbial communities: the riboswitch approach. *Clinical*
562   *microbiology and infection : the official publication of the European Society of Clinical*
563   *Microbiology and Infectious Diseases* 18 Suppl 4:35–9. DOI: 10.1111/j.1469-
564   0691.2012.03864.x.

565   Gutiérrez-Preciado A., Torres AG., Merino E., Bonomi HR., Goldbaum FA., García-Angulo VA.

566   2015. Extensive Identification of Bacterial Riboflavin Transporters and Their Distribution
567   across Bacterial Species. *Plos One* 10:e0126124. DOI: 10.1371/journal.pone.0126124.

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

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

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

576   Jakubauskas A., Kriukiene E., Trinkunaite L., Sapranauskas R., Jurenaite-Urbanaviciene S.,
577   Lubys A. 2009. Bioinformatic and partial functional analysis of pEspA and pEspB, two
578   plasmids from Exiguobacterium arabatum sp. nov. RFL1109. *Plasmid* 61:52–64. DOI:
579   10.1016/j.plasmid.2008.09.004.

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

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

587   Kim IG., Lee MH., Jung SY., Song JJ., Oh TK., Yoon JH. 2005. Exiguobacterium aestuarii sp.
588   nov. and Exiguobacterium marinum sp. nov., isolated from a tidal flat of the Yellow Sea in
589   Korea. *International Journal of Systematic and Evolutionary Microbiology* 55:885–889.
590   DOI: 10.1099/ijs.0.63308-0.

591   Krzywinski M., Schein J., Birol I., Connors J., Gascoyne R., Horsman D., Jones SJ., Marra MA.
592   2009. Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–
593   1645.

594   Lefebure T., Stanhope MJ. 2007. Evolution of the core and pan-genome of Streptococcus:
595   positive selection, recombination, and genome composition. *Genome Biol* 8:R71.

596   Li L., Jr CJS., Roos DS. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic
597   Genomes -- Li et al. 13 (9): 2178 -- Genome Research. :2178–2189. DOI:
598   10.1101/gr.1224503.candidates.

599   López M a., Zavala-Díaz de la Serna FJ., Jan-Roblero J., Romero JM., Hernández-Rodríguez C.
600   2006. Phylogenetic analysis of a biofilm bacterial population in a water pipeline in the Gulf
601   of Mexico. *FEMS microbiology ecology* 58:145–54. DOI: 10.1111/j.1574-
602   6941.2006.00137.x.

603   Merino E., Jensen R a., Yanofsky C. 2008. Evolution of bacterial trp operons and their

604   regulation. *Current Opinion in Microbiology* 11:78–86. DOI: 10.1016/j.mib.2008.02.005.

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

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

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

613   Nawrocki EP., Kolbe DL., Eddy SR. 2009. Infernal 1 . 0 : inference of RNA alignments.
614   25:1335–1337. DOI: 10.1093/bioinformatics/btp157.

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

617   Ordoñez OF., Flores MR., Dib JR., Paz A., Farías ME. 2009. Extremophile culture collection
618   from Andean lakes: extreme pristine environments that host a wide diversity of
619   microorganisms with tolerance to UV radiation. *Microbial Ecology* 58:461–473. DOI:
620   10.1007/s00248-009-9527-7.

621   Ordoñez OF., Lanzarotti E., Kurth D., Gorriti MF., Revale S., Cortez N., Vazquez MP., Farias
622   ME., Turjanski AG., Farías ME., Turjanski AG. 2013. Draft Genome Sequence of the
623   Polyextremophilic Exiguobacterium sp . Strain S17, Isolated from Hyperarsenic Lakes in
624   the Argentinian Puna. *Genome Announcements* 1:e00480-13. DOI: 10.1186/1471-2164-9-
625   75.Rodrigues.

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

628   Padan E., Bibi E., Ito M., Krulwich TA. 2005. Alkaline pH homeostasis in bacteria: New
629   insights. 1717:67–88. DOI: 10.1016/j.bbamem.2005.09.010.

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

633   Quevillon E., Silventoinen V., Pillai S., Harte N., Mulder N., Apweiler R., Lopez R. 2005.
634   InterProScan: Protein domains identifier. *Nucleic Acids Research* 33:116–120. DOI:
635   10.1093/nar/gki442.

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

638   Rasko DA., Rosovitz MJ., Myers GS., Mongodin EF., Fricke WF., Gajer P., Crabtree J.,
639   Sebaihia M., Thomson NR., Chaudhuri R., Henderson IR., Sperandio V., Ravel J. 2008.
640   The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli

641   commensal and pathogenic isolates. *J Bacteriol* 190:6881–6893.

642   Rebollar E a., Avitia M., Eguiarte LE., González-González A., Mora L., Bonilla-Rosso G.,
643   Souza V. 2012. Water-sediment niche differentiation in ancient marine lineages of
644   Exiguobacterium endemic to the Cuatro Cienegas Basin. *Environmental Microbiology*
645   14:2323–33. DOI: 10.1111/j.1462-2920.2012.02784.x.

646   Rissman AI., Mau B., Biehl BS., Darling AE., Glasner JD., Perna NT. 2009. Reordering contigs
647   of draft genomes using the Mauve Aligner. *Bioinformatics* 25:2071–2073. DOI:
648   10.1093/bioinformatics/btp356.

649   Rodrigues DF., Ivanova N., He Z., Huebner M., Zhou J., Tiedje JM. 2008. Architecture of
650   thermal adaptation in an Exiguobacterium sibiricum strain isolated from 3 million year old
651   permafrost: a genome and transcriptome approach. *BMC genomics* 9:547. DOI:
652   10.1186/1471-2164-9-547.

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

655   Ronquist F., Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed
656   models. *Bioinformatics* 19:1572–1574. DOI: 10.1093/bioinformatics/btg180.

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

658   Rouli L., Merhej V., Fournier P-E., Raoult D. 2015. The bacterial pangenome as a new tool for
659   analyzing pathogenic bacteria. *New Microbes New Infect* 7:72–85. DOI:
660   10.1016/j.nmni.2015.06.005.

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

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

666   Schöniger M., Von Haeseler A. 1994. A stochastic model for the evolution of autocorrelated
667   DNA sequences. *Molecular Phylogenetics and Evolution* 3:240–247. DOI:
668   http://dx.doi.org/10.1006/mpev.1994.1026.

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

671   Silver S., Phung LT. 2005. A bacterial view of the periodic table: genes and proteins for toxic
672   inorganic ions. *Journal of industrial microbiology & biotechnology* 32:587–605. DOI:
673   10.1007/s10295-005-0019-6.

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

676   Ventosa A., Nieto JJ., Oren A. 1998. Biology of Moderately Halophilic Aerobic Bacteria.

677   *Microbiology and Molecular Biology Reviews* 62:504–544.

678   Vishnivetskaya T a., Kathariou S. 2005. Putative transposases conserved in Exiguobacterium
679   isolates from ancient Siberian permafrost and from contemporary surface habitats. *Applied*
680   *and Environmental Microbiology* 71:6954–6962. DOI: 10.1128/AEM.71.11.6954-
681   6962.2005.

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

684   Vishnivetskaya T a., Lucas S., Copeland A., Lapidus A., del Rio TG., Dalin E., Tice H., Bruce
685   DC., Goodwin L a., Pitluck S., Saunders E., Brettin T., Detter C., Han C., Larimer F., Land
686   ML., Hauser LJ., Kyrpides NC., Ovchinnikova G., Kathariou S., Ramaley RF., Rodrigues
687   DF., Hendrix C., Richardson P., Tiedje JM. 2011. Complete genome sequence of the
688   thermophilic bacterium Exiguobacterium sp. AT1b. *Journal of Bacteriology* 193:2880–
689   2881. DOI: 10.1128/JB.00303-11.

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

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

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

699   Woese CR., Kandler O., Wheelis ML. 1990. Towards a natural system of organisms: proposal
700   for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of*
701   *Sciences of the United States of America* 87:4576–4579. DOI: 10.1073/pnas.87.12.4576.

702   Yuan M., Chen M., Zhang W., Lu W., Wang J., Yang M., Zhao P., Tang R., Li X., Hao Y., Zhou
703   Z., Zhan Y., Yu H., Teng C., Yan Y., Ping S., Wang Y., Lin M. 2012. Genome sequence
704   and transcriptome analysis of the radioresistant bacterium deinococcus gobiensis: Insights
705   into the extreme environmental adaptations. *PLoS ONE* 7:1–11. DOI:
706   10.1371/journal.pone.0034458.

707   Zhang Y., Qiu S. 2015. Phylogenomic analysis of the genus Ralstonia based on 686 single-copy
708   genes. *Antonie van Leeuwenhoek*. DOI: 10.1007/s10482-015-0610-4.

709

Figure 1

Evolutionary history of the genus *Exiguobacterium*

A) **Phylogenetic reconstruction with the 16S rRNA gene using MrBayes** (see Methods). The values of those branches that are not supported by 100% MrBayes parametric bootstrap, are shown in black. The phylogeny is rooted with a Bacillus outgroup. The genus Exiguobacterium separates in two clades: I, dominated by psychrophilic and II, dominated by alkaphilic species with a marine origin. 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). SILVA Accession Numbers of the sequences used are as follows: *B. anthracis* str. Ames AE017334; *B. subtilis* str. 168, AL009126; *B. halodurans* str. C-125, BA000004; *O. iheyensis,* BA000028; *E. marinum*, AY594266; *E.* sp. AT1b, CP001615; *E. arabatum*, FM203124; *E.* sp. S17, FN994191; *E.* sp. 8-11-1, KC757126; *E. aurantiacum,* DQ019166; *E.* sp. GIC31, EU282458; *E. mexicanum,* JX094173; *E. pavilionensis* str. RW-2, ATCL01000001; *E. acetylicum,* DQ019167; *E. undae,* FN870071; *E. oxidotolerans*, AB105164; *E. antarcticum* str. B7, CP003063; *E. sibiricum* str. 255-15, CP001022. The sequence from *E.* sp. MH3 was retrieved from its genome (NCBI accession number: NC_022794).

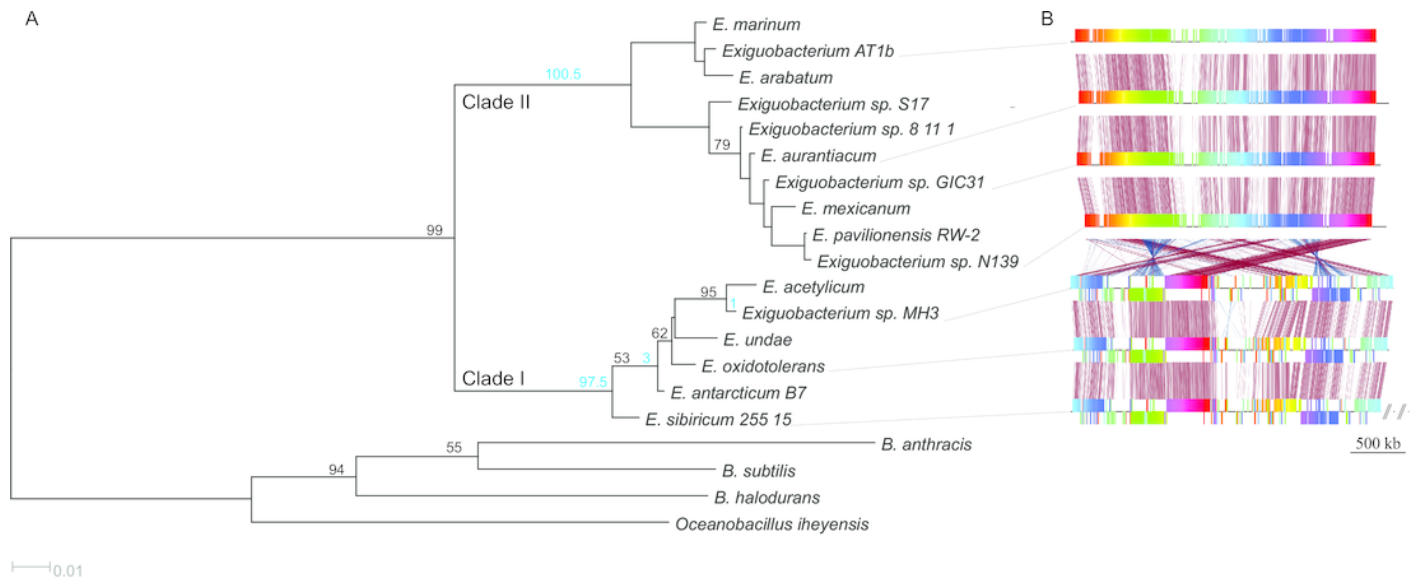
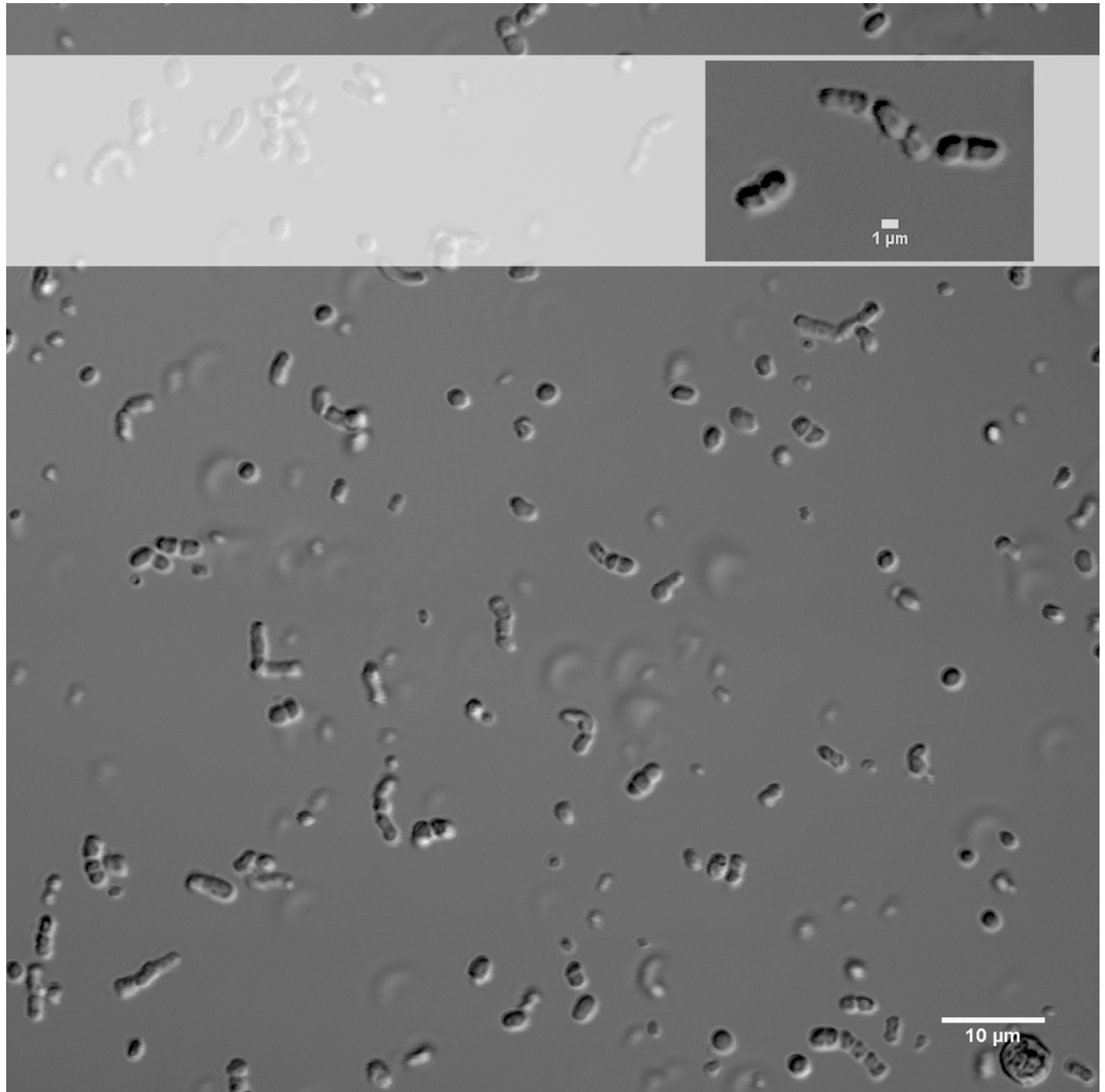
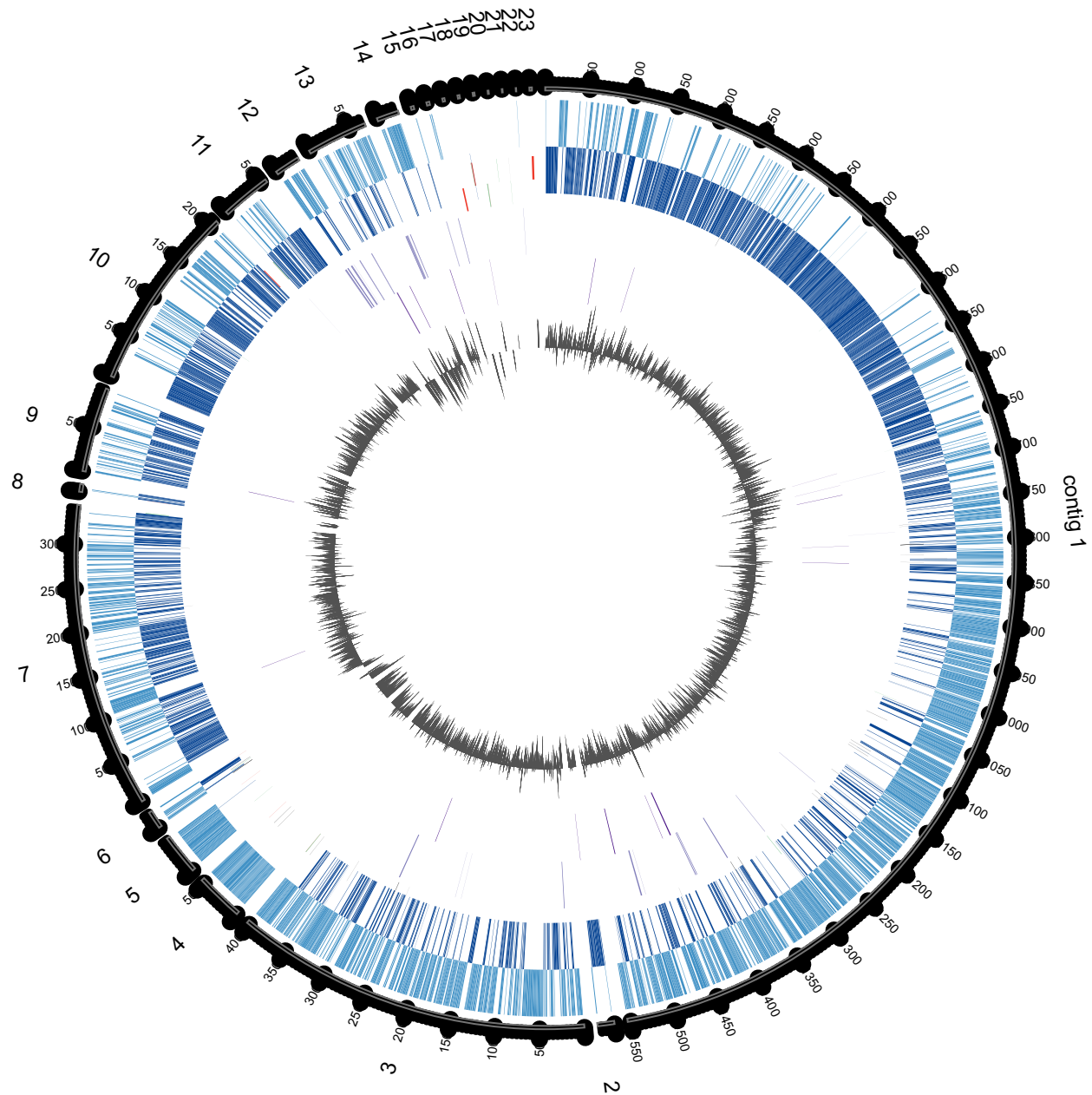



Figure 2

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.

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).

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

 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")

1   **Table 1.** Classification and general features of *Exiguobacterium pavilionensis* 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) TAS (De Vos et al., 2009; Vishnivetskaya, Kathariou & Tiedje, 2009)
	Genus Exiguobacterium	
	Species *Exiguobacterium pavilionensis*	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

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

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

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.

1   **Table 3**. Nucleotide content and gene count levels of the *E. pavilionensis* 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	

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

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.

1   **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

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