View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Bangor University Research Portal

PRIFYSGOL BANGOR / BANGOR UNIVERSITY

PRIFYSGOL
BANGOR
UNIVERSITY

Biology of archaea from a novel family Cuniculiplasmataceae (Thermoplasmata) ubiquitous in hyperacidic environments

Golyshina, Olga; Kublanov, Ilya V.; Hai, Tran; Korzhenkov, Alexei A.; Lunsdorf, Heinrich; Nechitaylo, Taras Y.; Toshchakov, Stepan V.; Golyshin, Peter

Scientific Reports

DOI:
10.1038/srep39034

Published: 01/12/2016

Peer reviewed version

Cyswllt i'r cyhoeddiad / Link to publication

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Golyshina, O., Kublanov, I. V., Hai, T., Korzhenkov, A. A., Lunsdorf, H., Nechitaylo, T. Y., Toshchakov, S. V., & Golyshin, P. (2016). Biology of archaea from a novel family Cuniculiplasmataceae (Thermoplasmata) ubiquitous in hyperacidic environments. *Scientific Reports*, *6*, [39034]. https://doi.org/10.1038/srep39034

Hawliau Cyffredinol / General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

11. May. 2021

1 **Biology of archaea from a novel family *Cuniculiplasmataceae***
2 ***(Thermoplasmata)* ubiquitous in hyperacidic environments**
3
4
5 **Olga V. Golyshina[1]*, Ilya V. Kublanov[2], Hai Tran[1], Alexei A. Korzhenkov[3],**
6 **Heinrich Lünsdorf[4], Taras Y. Nechitaylo[5], Sergey N. Gavrilov[2], Stepan V.**
7 **Toshchakov[3] and Peter N. Golyshin[1]**
8
9 [1] School of Biological Sciences, Bangor University, Deiniol Rd, Bangor, LL57 2UW,
10 UK
11 [2] Winogradsky Institute of Microbiology, Research Center for Biotechnology
12 Russian Academy of Sciences, Prospect 60-Letiya Oktyabrya 7/2, Moscow,
13 117312, Russia
14 [3] Immanuel Kant Baltic Federal University, 236040 Kaliningrad, Russia
15 [4] Central Unit of Microscopy, Helmholtz Centre for Infection Research,
16 Inhoffenstrasse 7, Braunschweig, 38124, Germany
17 [5] Insect Symbiosis Research Group, Max Planck Institute for Chemical Ecology,
18 Hans-Knöll-Strasse 8, Jena, 07745, Germany.
19
20
21 *Corresponding author Tel.: +44 1248 383629; Fax: +44 1248 382569; e-mail:
22 o.golyshina@bangor.ac.uk
23
24
25 **Key words: *Cuniculiplasma*, *Cuniculiplasmataceae*, G-plasma,**
26 ***Thermoplasmatales*, acidophilic archaea, acid mine drainage (AMD) systems**

27 Project accession numbers at EMBL-EBI European Nucleotide Archive (ENA):
28 PRJEB12275 (for the strain PM4)
29 PRJEB12276 (for the strain S5)
30

31 **ABSTRACT**

32 The order *Thermoplasmatales* (*Euryarchaeota*) is represented by the most
33 acidophilic organisms known so far that are poorly amendable to cultivation. Earlier
34 culture-independent studies in Iron Mountain (California) pointed at an abundant
35 archaeal group, dubbed ‘G-plasma’. We examined the genomes and physiology of
36 two cultured representatives of a Family *Cuniculiplasmataceae,* recently isolated
37 from acidic (pH 1-1.5) sites in Spain and UK that are 16S rRNA gene sequence-
38 identical with ‘G-plasma’.
39 Organisms had largest genomes among *Thermoplasmatales* (1.87-1.94 Mbp), that
40 shared 98.7-98.8% average nucleotide identities between themselves and ‘G-
41 plasma’ and exhibited a high genome conservation even within their genomic
42 islands, despite their remote geographical localisations. Facultatively anaerobic
43 heterotrophs, they possess an ancestral form of A-type terminal oxygen reductase
44 from a distinct parental clade. The lack of complete pathways for biosynthesis of
45 histidine, valine, leucine, isoleucine, lysine and proline pre-determines the reliance
46 on external sources of amino acids and hence the lifestyle of these organisms as
47 scavengers of proteinaceous compounds from surrounding microbial community
48 members. In contrast to earlier metagenomics-based assumptions, isolates were
49 S-layer-deficient, non-motile, non-methylotrophic and devoid of iron-oxidation
50 despite the abundance of methylotrophy substrates and ferrous iron *in situ*, which
51 underlines the essentiality of experimental validation of bioinformatic predictions.

52

53 **Introduction**
54

55 Acidic environments are widely distributed across the globe and are represented
56 by natural (e.g. volcanic or geothermally heated), or man-made (mines or acid
57 mine drainage (AMD)) sites, with a constantly low pH[1]. Microbial communities
58 inhabiting such niches were considered to be of a relatively low complexity[2],
59 however, recent OMICS studies pointed at a greater variety of yet uncultured
60 prokaryotes[1]. Due to the low numbers of cultured microorganisms that may serve
61 as a functional reference, their physiological features and hence the roles in the
62 environment largely remain at the level of *in silico* predictions from metagenomic
63 data. In that context, while a certain success has been achieved in isolation of new
64 bacterial taxa from these specific environments[3], only a handful of cultured,
65 taxonomically described and physiologically studied archaeal representatives have
66 been obtained[3]. Recent data based on genomes assembled from metagenomes
67 documented a number of archaeal clades mostly affiliated with the order
68 *Thermoplasmatales,* phylum *Euryarchaeota*[4]. Among archaeal populations from
69 the above order, cultured members of *Ferroplasmaceae* together with yet
70 uncultured archaea from the so-called ‘alphabet plasmas’ were the most abundant
71 and hence suggested to play important roles in carbon cycling in the environment[5].
72 Initially identified in 16S rRNA gene clone libraries from Iron Mountain[6], these
73 archaea have later been found in a number of acidic environments of different
74 temperature regimes[1]. Their presence in iron-rich environments have quite logically
75 promoted discussions on their iron oxidation potential. Apart from the iron oxidation
76 experimentally confirmed only in cultured members of *Ferroplasmaceae*[7,8], other
77 *Thermoplasmatales* were described as facultatively anaerobic heterotrophs[9]. Their

78 appearance in biofilms alongside with chemolithoautotrophs suggests that the
79 metabolism of this group may be depended on organic compounds (sugar
80 polymers/oligomers, peptides, lipids or carbohydrate monomers) derived from
81 primary producing organisms[10]. “Alphabet plasmas” were furthermore predicted to
82 oxidise carbon monoxide and utilise methylated compounds[4]. However, the dearth
83 of experimental evidence has largely limited our entire current understanding on
84 metabolism, physiology, and environmental roles of these archaeal lineages.
85 One of important members of *Thermoplasmatales* in AMD systems from Iron
86 Mountain (California) was a group of organisms dubbed ‘G-plasma’, which was so
87 abundant, that the genomes of some of the representatives were almost fully
88 assembled[2,4]. These organisms were the third-abundant community members
89 (following *Leptospirillum* spp. “Group II and III”) and contributed up to 22 % of total
90 community proteome[11]. Elsewhere, ‘G-plasma’ contributed approx. 15 % of total
91 metagenomic reads in this environment[12]. However, despite their abundance and
92 ubiquity these organisms escaped the cultivation until very recently, when first
93 representatives were isolated from Cantereras AMD site (Spain) and Parys
94 Mountain/Mynydd Parys (Anglesey, UK) and described as representatives of a
95 novel family, *Cuniculiplasmataceae,* new genus and species *Cuniculiplasma*
96 *divulgatum* within the order *Thermoplasmatales*[13]. The 16S rRNA gene sequences
97 of isolates PM4 and S5 were identical to those from ‘G-plasma’ cluster from
98 metagenomic data analysed at Richmond mine at Iron Mountain, USA[2,4], terrestrial
99 acidic springs in Japan[14], high-temperature fumarole and acidic biofilm from
100 Mexico (GenBank Acc Nrs. JX997948, AB6000334 and KJ907759), Frasassi
101 hydrogen sulphide-rich cave system, Italy[12] and from AMD system, Los Rueldos,
102 Spain[1,10] and other low pH systems. Altogether, these documentations point at the

103 ubiquity of *Cuniculiplasmataceae*-related organisms in acidic systems and volcanic
104 areas (Fig. 1, Supplementary Table S1).
105 New isolates provided a great opportunity to perform for a first time the
106 comparative genomic analysis of very closely related members of *Thermoplasmata*
107 from very distant geographic locations, to analyse their physiology and functions
108 related to the environment in the context of the earlier genomic predictions, and,
109 finally, to analyse their evolutionary relationships with other clades within the class
110 *Thermoplasmata*, which harbours organisms known as acidophilic 'champions'[9,15].
111

112 Results

113 Physiological traits: *in silico* predictions in 'G-plasma' vs experimental data

114 *Iron oxidation.*
115 Despite earlier suggestions of iron-oxidising capabilities based on the occurrence
116 of rusticyanin/sulfocyanin-encoding gene homologs[4], no iron oxidation was
117 confirmed with ferrous sulfate and pyrite in either *C. divulgatum* isolate.
118 Noteworthy, the presence of genes for rusticyanin/sulfocyanin homologs might not
119 necessarily be connected with the iron oxidation in archaea of the order
120 *Thermoplasmatales,* e.g. *Picrophilus torridus* does not oxidise ferrous iron despite
121 the presence of sulfocyanin. It was suggested[16] that this respiratory complex in *P.*
122 *torridus* is situated on a genomic island, which seems also to be the case for
123 rusticyanin/sulfocyanin genes acquired by a lateral transfer in *C. divulgatum* S5 (s.
124 below) and *F. acidiphilum* (Golyshina *et al.*, in preparation).
125

126 *Archaeal flagella and pili*

127 In ‘G-plasma’, the full operon encoding FlaBCDEFGHIJ with individual proteins
128 being homologous to those from *Methanococcus voltae* and *Halobacterium*
129 *salinarum* has been reported earlier[4], however, corresponding loci have not been
130 detected in either genome of *Cuniculiplasma divulgatum*. Our analysis suggested
131 that *M. voltae* and *H. salinarum* flagellar proteins do not have significant (e-values
132 <0.01 and query coverage > 50%) BLAST hits in ‘G-plasma’ *in silico* translated
133 proteome. Electron microscopy of *C. divulgatum* grown under optimal condtions did
134 not provide evidence for an archaellum, but occasionally showed the presence of
135 distinct pili[13], (s. also Fig. 2c). This feature is also reflected in the genomic data of
136 both isolates of *Cuniculiplasma*, as discussed further (s. subsection ‘*Secretion*
137 *system and motility’*).
138

139 Regarding the S-layer prediction in ‘G-plasma’, corresponding genes to be linked
140 with S-layer formation[4] were also found in both *Cuniculiplasma* genomes. These
141 genes annotated in ‘G-plasma’ to code for “S-layer protein *P. torridus*”[4] (and
142 annotated as a surface protein in *P. torridus* itself)*,* are affiliated with COG3391,
143 arCOG0652 and arCOG2560 that have five homologs in each *Cuniculiplasma*
144 *divulgatum* genome). Additionally, genes encoding oligosaccharyltransferase AglB
145 in ‘G-plasma’ are present in both genomes of *Cuniculiplasma* strains, as well.
146 However, as revealed by electron microscopy, the cells of strains S5 and PM4
147 were only surrounded by cytoplasmic membranes and lacked distinct (predicted in
148 ‘G-plasma‘) S-layers (Fig. 2 a, b). An S-layer should provide a certain rigidity to
149 cells and its absence is consistent with the characteristic pleomorphism in *C.*
150 *divulgatum,* as exemplified in Fig. 2 c,d. Apparently, within the order
151 *Thermoplasmatales,* the cell wall-deficient members clearly outnumber S-layer-

152 exhibiting organisms, which are represented only by *Picrophius* spp.[9]. This feature
153 is also reflected in the genomic data of two strains, as discussed further.
154
155 *Methylotrophy*
156 In the growth experiments performed with both strains of *C. divulgatum* with a
157 range of methylated compounds[13] we were not able to confirm the methylotrophy
158 earlier predicted in 'G-plasma'[4]. In this regard, the genes predicted to be present in
159 'G-plasma', namely methenyl tetrahydrofolate cyclohydrolase and formyl-
160 tetrahydrofolate synthetase have also been found in both *Cuniculiplasma*
161 genomes. However, the gene encoding 'methanol dehydrogenase' in 'G-plasma'
162 has not been confirmed in *Cuniculiplasma*. Furthermore, the protein referred as
163 such in 'G-plasma' itself had a low amino acid sequence identity (>26%) to alcohol
164 dehydrogeneases of unknown substrate specificity and was equally (dis)similar
165 with maleylacetate reductases. The very homolog was found in the S5 genome,
166 but not in PM4. Among tested substrates, e.g. methylamines, could not be utilised
167 by *Cuniculiplasma* isolates since no genes for methylamine dehydrogenase or
168 dimethylamine and trimethylamine dehydrogenase were found. Whatever the case,
169 methylotrophy was not experimentally confirmed in any *Thermoplasmatales,* even
170 though the methanol is a common product of organic matter degradation and may
171 be available in studied environments.
172
173 **Genome analysis of *Cuniculiplasmataceae***
174 *Genome statistics*
175 The genomes of *C. divulgatum* strains (Table 1) are larger as compared to the
176 relatives from *Thermoplasma* spp (1.58 Mbp for *T. volcanium* and 1.56 Mbp for *T.*

177 *acidophilum*) and *Picrophilus torridus* (1.55 Mbp), being within the common range
178 to archaea of the family *Ferroplasmaceae* (1.94 Mbp for *“Ferroplasma*
179 *acidarmanus”,* 1.75-1.78 Mb for *Acidiplasma aeolicum* and 1.74 Mbp for *A.*
180 *cupricumulans*). Low G+C contents of genomic DNA of strains S5 and PM4 are
181 rather typical for *Thermoplasmatales*[17].
182
183 *Genome comparisons*
184 The three genomes exhibited a high average nucleotide identity (ANI)[18] and
185 average amino acid identity (AAI)[19], which also supports similar physiological
186 patterns in both isolates: strains S5 and PM4 had 98.8 % ANI, while ANI of both
187 isolates with ‘G-plasma’ genome were about 98.7 and 98.4 %, respectively,
188 pointing at their similar evolutionary trajectories despite transcontinental
189 localisation of their niches and highly complementary microbial structure and gene
190 pools in AMD settings[1] (also s. Fig. S1 for the AAI data).
191 The core *in silico* proteome of *C. divulgatum* strains and ‘G-plasma’ is represented
192 by 1174 protein groups. 111 protein clusters were identified as exclusively
193 distributed among PM4 and S5 strains, 13 among PM4 and ‘G-plasma’ and 27
194 among S5 and ‘G-plasma’ (Fig. 3, 4 and Fig. S1, Supplementary Table S2). 79, 52
195 and 114 unique single-copy genes and 1, 1 and 10 strain-specific paralogue
196 clusters were identified for S5, PM4 and ‘G-plasma’ respectively. Analysis of their
197 distribution across the chromosomes revealed that most of them are highly
198 clustered (Fig. 4), supporting the hypothesis that LGT (lateral gene transfer) is an
199 important driving force in evolution of AMD-related microorganisms[20], however with
200 very similar patterns of foreign DNA integration in the genomes of recipients.
201

202 *Lateral gene transfer (LGT), genomic islands (GIs) and defence systems*

203 Analysis of arCOG distribution within variable and core parts of *Cuniculiplasma*-
204 related *in silico* proteomes revealed a significant enrichment in “Defense
205 mechanisms” group in PM4 strain. This observation together with the fact that PM4
206 possesses 92 non-redundant CRISPR spacers as opposed to 52 in S5 strain and
207 only 10 in ‘G-plasma’ give an opportunity to speculate that Parys Mountain/Mynydd
208 Parys mine is characterised by much higher viral load than other investigated acid
209 mine habitats[21]. In turn, unique and accessory part of ‘G-plasma’ genome
210 characterised by the lowest proportion of ‘defence mechanisms’ is highly enriched
211 with ‘replication, recombination and repair’ proteins including integrases,
212 transposases and recombinases pointing on higher level of genome mobility in ‘G-
213 plasma’ (Fig. 3). Another point related to arCOG distribution is the prevailing
214 comparative number of unique strain-specific proteins in S5 for categories energy
215 production and conversion, cell cycle control, transcription, inorganic ions transport
216 and metabolism (Fig. 3).

217 Lateral gene transfer (LGT), genomic islands (GIs) and defence systems.

218 The strain S5 harboured ten GIs in its genome, whereas its counterpart from
219 Parys Mt/Mynydd Parys only four (Fig. 4 (a, b) and Supplementary Table S2). As
220 expected, numerous insertion sequences elements (IS), integrases and
221 transposases from different families (IS3, IS5, IS6, IS66, IS256, IS200/605, IS110,
222 IS1634) were associated with the GIs, as well as tRNAs reflecting the commonality
223 of tRNA co-occurrence in genomic islands[22]. The G+C molar content in predicted
224 GIs varied within the range 37.7 – 43.2 %, i.e. marginally higher than average
225 values in PM4 and S5 genomes (Supplementary Table S3), which may be a result
226 of old integration events and consequent DNA amelioration in GIs making GC

227 similar to that in the core genomes. Notably, a slight difference between G+C-
228 content in genomes of S5 and PM4 strains (37.16% in PM4 vs 37.30 in S5) is
229 determined by the presence of six additional GIs in the former isolate. Analysis of
230 taxonomic affiliation of GIs revealed that almost all lateral transfers originated from
231 other acidophilic euryarchaea. This observation implies the existence of a highly
232 mobile gene pool in acidophilic *Archaea*, which determines rapid adaptations of
233 *Thermoplasmatales* members to toxic concentration of heavy metals and to a high
234 viral load.

235 Thus, some GIs could clearly be attributed to 'defence' islands, e.g. GI3, GI7 and
236 GI9-10 in S5 and GI4 in PM4 due to the localisation therein of genes for restriction-
237 modification and toxin-antitoxin systems. Others (e.g. GI 4, GI 5 and GI 8 from S5)
238 were transport-, efflux-, metal- and oxidative stress response-related). GI1 from
239 the strain S5, which is absent in PM4, harboured an array of genes for site-specific
240 recombinases, metal-transporting ATPases, multipass membrane proteins,
241 metallochaperones, cupredoxin COX2 family proteins, heavy metal reductases,
242 and rustycyanin/sulfocyanin homolog.

243 We have identified several toxin-antitoxin systems (TAS) -encoding genes, mostly
244 associated with GIs in both isolates. The most abundant ones were represented by
245 *vap*BC of the type II system: six clusters of corresponding ORFs in PM4, and
246 seven in S5 and, besides three *vap*B toxin genes were found across chromosomes
247 in both *Cuniculiplasma* isolates. In addition, three clusters of genes were found in
248 PM4 and two such loci in S5 with corresponding MazE and MazF family proteins
249 affiliated with COG2336/arCOG03943 and COG2337, respectively**.**

250 Furthermore, three and two *rel*EF loci were identified in PM4 and S5 genomes,
251 correspondingly. Commonly, TAS are known to be stress response-connected and

252 lateral gene transfer-related[23,24], which is confirmed by the GI analysis. Notably, no
253 TAS were previously reported in *Thermoplasmatales*[25].
254 All genomes of *Cuniculiplasma* spp. showed the presence of Clustered Regularly
255 Interspaced Short Palindromic Repeats (CRISPR)-Cas defence systems: in S5, we
256 have identified the cluster of genes for Cas3, Csx17, Cas7, Cas5, Cas4/Cas1 and
257 Cas2 with an adjacent CRISPR repeat region with 57 spacers. Interestingly, all
258 proteins exhibited 100% polypeptide identity with counterparts from 'G-plasma'
259 (apart from Cas4 and 1 which had psi-blast hits of about 54% identity with
260 acidobacterial polypeptides).
261 ATDV01000019 contig of 'G-plasma' exhibited a remarkable similarity in gene
262 arrangement (ADMU5_GPLC00019G0101-G0107) with the corresponding region
263 in S5 (CIP_1636-1642) albeit with only 10 spacers of repeats found on the
264 terminus of the contig ATDV01000011. According to[26] systems from 'G-plasma' and
265 S5 can be classified as Type I-C. The strain PM4, in contrast to the above, coded,
266 in this order, for Cas6 endoribonuclease, Cas8b, Cas7, Cas5, Cas3, Cas4, Cas1
267 and Cas2, flanked by a repeats-spacers array of 92 spacers, suggesting its
268 affiliation with the Type I-B system[26]. Interestingly, all sequences of Cas proteins
269 were equally distant (28-57% sequence identity (Supplementary Table S4) with the
270 proteins from '*F. acidarmanus*' and other archaea and, to the same extent, with
271 polypeptides from representatives of *Bacteria,* e.g*. ∂-Proteobacteria* or
272 *Acidobacteria* (pointing at an unclear origin of corresponding gene clusters).
273 Remarkably, very similar pseudogenes CPM_1008 and CSP5_0996 for Cas1 were
274 detected in both isolates, in similar locations on chromosomes, within the same
275 genomic context in the region severely affected by transposon integration and
276 pseudogenisation. Analysis of CRISPR repeats in S5, PM4 and 'G-plasma' by

277 blastn-short algorithm revealed no cross-matches of spacers between these three
278 genomes. Nevertheless, eight of 92 PM4 repeats and four of 57 S5 spacers
279 showed high (90-100%) identity with sequences of Richmond mine microbial and
280 viral communities[21,27], suggesting the existence of viruses common for these
281 extreme acidic ecosystems. Interestingly, CRISPR array of PM4 contains two
282 spacers with the significant level of similarity (83 and 96%) to marine metagenomic
283 sequences (Supplementary Table S5). Despite a significantly high probability of
284 false positive hits (e-values are 0.015 and 0.14, respectively), this finding might be
285 speculated as relic genomic signatures of an ancient hydrothermal ecosystem
286 which existed 480-360 my BP in the place of contemporary Parys Mountain site[28].
287 From the analysis of GIs in *Cuniculiplasma* spp. two important facts become
288 apparent. First, the co-occurrence of GIs and the majority of 'unique' genes
289 (numbers in the outermost segments in Fig. S1 and green lines in Fig. 3). Most
290 'unique' genes had likely been acquired from organisms other than
291 *Thermoplasmata* and had no hits above the e-value cut-off (0.005) either with
292 'alphabet plasmas' or with isolates from cultured/genome-sequenced
293 *Thermoplasmata*, suggesting a high probability of lateral gene transfer also in the
294 vicinity of GIs. Second, a remarkable similarity in gene arrangements was
295 observed within some GIs in both strains and their positioning in both
296 chromosomes, i.e. in 'defence islands' GI9-10 of S5 and GI4 of PM4 (homologous
297 to 'G-plasma' contig ADMU5_GPLC00019G0004-G0013) (Supplementary Fig. S2
298 (b) and Supplementary Table S6) and GI2 of S5 and GI1 and 2 from PM4 that were
299 mostly composed by ORFs for hypothetical proteins conserved in both organisms
300 (Supplementary Fig. S2 (a)). Such conservation in gene arrangements in GIs is
301 indicative for an important role these genes may play in metabolism in iron-rich

302 environments and that they can relatively easily be transferred between organisms
303 and remain in genomes due to the selective pressure, providing a competitive
304 advantage, much like ‘catabolic transposons’ for xenobiotics or hydrocarbon
305 metabolism[29]. This was the case in, e.g., three transposases-adjacent operons in
306 strain S5 encoding metallochaperone and metalloreductases that showed high
307 similarities with counterparts in all *Thermoplasmatales* type strains.
308

309 *Secretion systems and motility*

310 In the genomes of both strains PM4 and S5 no operon essential for archaella
311 biogenesis (*fla*CDFGB)[30] was found, and consistently, no archaella and no motility
312 were observed by microscopy, despite earlier suggestions[4,13]. The strain PM4
313 exhibited filaments or pili-like cell surface structures[13], according with the presence
314 in genomes of genes for proteins of type IV pili biosynthesis. In accordance with
315 the recent census of archaeal clusters of orthologous groups of proteins (arCOG)
316 related with pili formation[31], we have identified principal components in both
317 genomes as follows. In S5, CSP5_0712 and CSP5_0715 encoded Type II
318 secretion system ATPase subunits (FlaI, arCOG01817) forming a gene cluster with
319 genes for CSP5_0713-14, encoding homologs of flagellar assembly proteins J2
320 and J1 (TadC, arCOG01808) and major pilins (FlaB/FlaF/PilA family, arCOG02423)
321 coded by clustered CSP5_1254-1255 and stand-alone CSP5_0804 and
322 CSP5_0881. The arrangement of two gene clusters harbouring six former gene loci
323 resembled that in both genomes of “*Aciduliprofundum*” spp.[31]. In the strain PM4,
324 corresponding loci were CPM_0710 and 0713 (secretion ATPases), CPM_0711-
325 0712 (TadC-like proteins) and CPM_1256-57, 0800 and 0878 (major pilins), with
326 the very same arrangement of gene clusters across the chromosome, as in the

327 strain S5. Function of these surface formations could be various: surface adhesion,
328 intercellular connection, DNA exchange or probably attachment to the substratum
329 rather than the motility[32]. Both strains encode type IV secretion components:
330 TraG/TraD/VirD4 family ATPases (arCOG04816) by CSP5_0791 and CPM_0795;
331 membrane protein (arCOG05340), VirB4 component (arCOG04034), multipass
332 protein (arCOG05369) and membrane protein (conserved in *Thermoplasmatales*
333 only) with four latter encoded by gene clusters CSP5_1185-1189 and CPM_1190-
334 93. Furthermore, both genomes encode Sec translocon components, preprotein
335 translocase subunits SecYE and Sec61beta, signal peptide peptidase and signal
336 recognition particle subunits and receptors. Another feature to be addressed here
337 is the presence of Sec-independent Tat pathway genes for folded proteins'
338 secretion. Twin-arginine translocase subunits A and C are presented in PM4 and
339 S5 genomes, which may be functional in an analogy with a Gram-positive bacterial
340 Tat system, known to work without additional TatB protein[33].
341
342 *Peptidases, peptide/amino acids transporters*
343 Consistently with the substrate preferences for proteinaceous compounds, each
344 genome contained more than 50 various peptidases. Among those, eight were
345 predicted to be secreted due to the presence of signal peptides. Five peptidases
346 were most probably responsible for extracellular hydrolysis of proteins and
347 peptides: three serine peptidases of S53 family and two thermopsins, aspartic
348 peptidases of A5 family[34]. S53 family peptidases have 3D structures similar with
349 other representatives of SB clan, their distant homologs, subtilases of S8 family,
350 but differ in acidic pH optima for activity. Since all *Thermoplasmatales* are
351 extremely acidophilic microorganisms, it is quite logic that S8 peptidases-coding

352 genes were not found in their genomes, and were ‘replaced’ by S53 peptidases. A
353 thermopsin, also characterised as an acidic endopeptidase[35] is another reflection
354 of adaptation of *Cuniculiplasma* spp. to extremely acidic conditions. The genomes
355 of the strains S5 and PM4 encoded two almost identical thermopsins, however,
356 one of S5 thermopsins lacked 130 amino acid on its N-terminus and hence lacking
357 secretion system motifs. Genomic context analysis revealed the presence of
358 various transporters in close vicinity of A5 peptidases of both strains and almost no
359 transporters in S53 neighbourhood. Among transporters, surrounding thermopsins,
360 the most probable amino acid and peptides importers were among the members of
361 Major Facilitator Superfamily (MFS, 2.A.1), according to TCDB database[36].
362
363 *TCA*
364 All genes, coding for TCA proteins were clearly identified in *Cuniculiplasma*
365 genomes except 2-oxoglutarate dehydrogenase (EC 1.2.4.2) and fumarate
366 reductase (EC 1.3.5.1). A 2-ketoacid dehydrogenase complex was found
367 (CSP5_0253-0256 and CPM_0219-0222), however it was related to rather 2-
368 oxoisocaproate dehydrogenase (EC 1.2.4.4) than to 2-oxoglutarate
369 dehydrogenase (EC 1.2.4.2) or pyruvate dehydrogenase (1.2.4.1). Still, the
370 conversion of 2-oxoglutarate to succinyl-CoA could be catalyzed by 2-oxoglutarate
371 synthases (CSP5_0284-0285 and CPM_0255-0253) and CSP5_1378-
372 1379/CPM_1377-1378). These ferredoxin-dependent enzymes are known to be
373 highly sensitive to oxygen, thus, presumably being active during anaerobic growth
374 of *C. divulgatum* or being highly stable to oxygen as it was shown for a homolog
375 from *Mycobacterium tuberculosis*[37]. CSP5_1895 and CPM_1834 (COG1027) are
376 homologous to several characterised class II aerobic fumarases (EC 4.2.1.2),

377 however the phylogenetic analysis shows (Supplementary Fig. S3) their marginally
378 closer relatedness with aspartases (EC 4.3.1.1) than with fumarases (yet with high
379 AA identity/similarity values (38/57%) with the Class II fumarase from *Sulfolobus*
380 sp.). Whatever the case, a possible absence of fumarase would imply
381 incompleteness of the TCA cycle, however it would still be able to generate the
382 proton motive force via the Complex II (succinate dehydrogenase CSP5_0486-
383 0489 and CPM_0468-0451). As expected, glyoxylate bypass seems to be
384 inoperative: isocitrate lyase was found, but not the malate synthase.
385 In the course of growth of *C. divulgatum* on peptides, the lack of recirculation of
386 TCA metabolites due to its incompleteness can be compensated by their synthesis
387 from amino acids. During potential sugars-driven growth, PEP can be converted to
388 oxaloacetate in a reversible reaction (which is not favourable, but possible),
389 catalysed by GTP-dependent phosphoenolpyruvate carboxykinase (CSP5_1337
390 and CPM_1336) while malate or oxaloacetate can be synthesized from pyruvate
391 by a reverse reaction catalysed by malic enzyme (CSP5_0838, CPM_0835).
392 Despite the generation of the proton motive force at aerobic growth (complex II) on
393 peptides or sugars (the latter was not confirmed experimentally in the current
394 experimental setup) TCA cycle enzymes play a crucial role in anabolism during
395 growth on peptides at both aerobic and anaerobic conditions. For example, the
396 mentioned above GTP-dependent phosphoenolpyruvate carboxykinase and malic
397 enzyme uses its metabolites for the first stages of gluconeogenesis:
398 phosphoenolpyruvate and pyruvate synthesis, respectively.
399
400 *NADH dehydrogenase*

401 Both *Cuniculiplasma* genomes contain genes for four major respiratory complexes,
402 with some unusual details, as specified below. A set of genes of the proton-
403 translocating type I NADH-dehydrogenase (complex I) *nuoABCDHIJJKLMN* is
404 encoded by CSP5_1737-1726 in the strain S5 and CPM_1708-1687 in the strain
405 PM4, in the same order. Both genomes encode neither NuoG subunit, nor subunits
406 NuoE or NuoF homologs essential to provide the catalytic site for NADH oxidation,
407 which raises doubts in NADH-oxidizing activity of the complex I and its involvement
408 in respiratory electron transfer chain in *C. divulgatum*. Alternative pathway of
409 electron inflow into the respiratory chain could be provided by succinate
410 dehydrogenase/fumarate reductase (Complex II), encoded by CSP5_0486-0489 in
411 S5 and CPM_0458-0461 in PM4 genomes. It should be mentioned that none of
412 *Thermoplasmatales* genomes available to date contain genes of NuoEF subunits,
413 indicating possibly inherent feature of respiratory complex I in the organisms of this
414 deep phylogenetic lineage and possible existence of other yet unknown alternative
415 mechanisms of electron flow from NADH oxidation – in analogy to those proposed
416 in aerobically respiring *Helicobacter pylori,* also lacking NuoEF subunits of the
417 complex I[38].
418

419 *Quinol oxidising complex III and oxygen respiration*

420 Quinol oxidising complex III in both *C. divulgatum* genomes is represented by
421 clustered genes of Rieske Fe-S protein and cytochrome *b* subunit of a typical
422 cytochrome bc_1 complex encoded by CSP5_1460-1459 in the S5 and CPM_1454-
423 1453 in the PM4 strains. These clusters are located remotely with the genes of
424 terminal respiratory reductases in both genomes. No genes of an alternative
425 complex III have been detected in *C. divulgatum* genomes.

426 Terminal oxygen reductases are represented in both *C. divulgatum* strains by a
427 typical cytochrome *bd* quinol oxidase (CSP5_0552-0553 and CPM_0524-0525)
428 and a heme copper oxygen reductase (HCO) encoded in the clusters CSP5_1313-
429 1312 and CPM_1312-1311. The first enzyme complex possesses a high affinity to
430 oxygen and is usually involved in oxygen detoxification or respiration under
431 microaerophilic conditions providing relatively low energy yield to the cell[39]. The
432 heme-copper oxygen reductase (complex IV) is a typical terminal enzyme of
433 aerobic respiratory electron transfer chain, coupling oxygen reduction to proton
434 translocation at aerobic or microaerophilic conditions. Sequence analysis of
435 catalytic subunits I of the heme-copper oxidases of *C. divulgatum* strains with a
436 web-based classifying tool (http://evocell.org)[40] clearly showed that both of them
437 belong to the type A1 oxygen reductases possessing two proton translocating
438 channels, and consequently, the highest proton pumping stoichiometry of $2H^+$ per
439 one electron[41,42,43]. Our phylogenetic reconstruction of full-size CoxI available so
440 far (Fig. 5) generally reproduced the recently reported topology of HCO
441 phylogenetic trees and revealed that the A1-type heme-copper oxidases of
442 *Cuniculiplasma* species form a distinct clade, most closely branching to B-type
443 oxygen reductases and to the root of all the other A-type reductases. Interestingly,
444 heme-copper oxygen reductases from other *Thermoplasmatales* (from *Acidiplasma*
445 and *Picrophilus* species) are located on a distinct clade of A-type oxidases (Fig. 5).
446 Furthermore, both *C. divulgatum* strains lack genes for membrane-integral oxygen
447 reductase subunits III and IV (either separately encoded or fused to the C-terminus
448 of the catalytic subunit I), while those were found in *Acidiplasma* and *Picrophilus*
449 genomes being fused with *coxI* genes. The subunits III and IV are regarded to be
450 distinguishing features of A-type (SoxM) heme-copper oxygen reductases acquired

451 during their evolution from less energetically effective and more ancient B-type
452 enzymes[42]. The lack of these subunits in *C. divulgatum* together with the
453 phylogenetic position of its CoxI proteins allows assuming that this organism
454 possesses an ancestral form of all known A-type terminal oxygen reductases.
455

456 A crucial point for the activity of the heme-copper oxygen reductase is the pathway
457 of electron transfer from the quinone pool or complex III. In *C. divulgatum* genomes,
458 there are no genes of type I monoheme *c*-type cytochromes, providing the electron
459 transfer from respiratory complexes III to terminal oxidases. Alternative pathway
460 could be driven by blue-copper redox proteins (cupredoxins), as described in several
461 acidophiles[44]. A homolog of such cupredoxins has been found to be involved in the
462 respiratory chain of *Ferroplasma acidiphilum*[45]. As mentioned above, the gene
463 encoding a cupredoxin rusticyanin was identified only in *C. divulgatum* strain S5
464 (CSP5_0076). The absence of both genes of type I cytotchrome *c* and
465 rusticyanin/sulfocyanin does not allow predicting the electron transfer pathway
466 between respiratory complexes III and IV in the strain PM4. The possibility still exists
467 that the complex IV in strain PM4 possesses quinol oxidizing activity and, similarly
468 to some other heme-copper oxidases, could directly accept electrons transferred
469 from the complexes I and II via the quinone pool. In such a case, the complex III in
470 strain PM4 would serve as an additional proton-translocating site, which is not
471 directly involved in oxygen respiration and could transfer electrons to an extrinsic,
472 yet unidentified acceptor. However, this assumption needs experimental evaluation.
473 All analysed genomes code for subunits K, E, C, F, A, B, D, H and I of V/A type H^+-
474 transporting ATP synthases, in this particular order (CSP5_0034-0042 and
475 CPM_0034-0042).

476 Further central metabolic and protein folding pathways detailed in SI suggest
477 *Cuniculiplasma* spp. largely share these with other *Thermoplasmata*.
478
479 ***Comparison with other Thermoplasmatales***
480 Phylogenomic analysis of *Thermoplasmata* based on concatenated amino acid
481 sequences of 11 conservative ribosomal proteins of each representative of the
482 phylum with a sequenced genome (Fig. 6 a), indicates a slightly different tree
483 topology than that suggested by 16S rRNA gene phylogenetic analysis[13], likely due
484 to this selection of particular molecular markers for phylogenetic reconstruction. On
485 the other hand, IMG COG-based hierarchical clustering placed
486 *Cuniculplasmataceae* representatives close to the root of the order
487 *Thermoplasmatales* (Fig. 6 b*).* This might be an indication that *Cuniculiplasma* spp.
488 share more parental properties than other cultivated members of
489 *Thermoplasmatales* and thus could be a good model for analysis of yet
490 uncultivated members of the class *Thermoplasmata*.
491 In contrast to other *Thermoplasmatales*, the genome of *C. divulgatum* strain PM4
492 (but not S5) had no restriction-modification system Type I. Pyrimidine and purine
493 conversion and utilization pathways, RNA processing and modification processes
494 showed their incompleteness in *C. divulgatum*, in comparison to the rest of
495 *Thermoplasmatales*. We also infer that amino acid biosynthesis category for other
496 *Thermoplasmatales* (*T. acidophilum, P. torridus, and "F. acidarmanus"*) showed
497 some discrepancies to *C. divulgatum*. Thus, *P. torridus* has been proposed to
498 possess all pathways for the amino acid synthesis[16]. *"F. acidarmanus"* occurred to
499 encode incomplete histidine, valine, leucine and isoleucine synthesis pathways[4].
500 The genomes inspection of *C. divulgatum* suggested, in addition to the above,

501 incomplete pathways for lysine and proline, pointing at the organisms' dependence
502 on external peptides and hence suggesting their role in the environment as
503 'scavengers'.
504 Incidentally, *C. divulgatum* and "*F. acidarmanus"* genomes, but not *T. acidophilum*
505 or *P. torridus* encode proteins for capsular heptose metabolism and
506 polyhydroxybutirate metabolism (with an exception of the gene encoding for
507 acetoacetyl-CoA synthetase (EC 6.2.1.16 in *"F. acidarmanus"*).
508 The organisms have a weak potential for synthesis of polymeric storage
509 compounds: both genomes for a similar folylpoly(gamma)glutamate synthase
510 (CPM_0655 and CPM_1446). The PM4 also contains an inorganic
511 polyphosphate/ATP-NAD kinase (CPM_0378) putatively active in energy gaining
512 from environmental polyphosphate deposits. However, no cell inclusions were
513 observed.
514 Formate dehydrogenase complex, involved into catabolism of C1 compounds, which
515 is a common trait for *T. acidophilum* and *"F. acidarmanus"* has not been verified in
516 either *Cuniciliplasma* genome. The gene coding for aquaporin Z (MIP superfamily)
517 was found in both genomes of *C. divulgatum*, potentially contributing to the osmotic
518 stress response and adaptive fitness, but absent in the genomes of other members
519 of *Thermoplasmatales*. Another distinctive feature is the lack of molybdenum
520 cofactor and coenzyme M biosynthesis in *C. divulgatum* genomes in contrary to
521 other *Thermoplasmatales*. Finally, the lack of ATP-dependent DNA ligases in *C.*
522 *divulgatum* genomes has been observed. The global analysis of distribution patterns
523 of arCOGs in *Thermoplasmatales* is further detailed in SI and suggests
524 *Cuniculiplasma* is a common member of the order.
525

526 **Discussion**

527 Isolation of previously uncultured microorganisms from the environment remains
528 one of the bottlenecks in microbiology hindering physiological and biochemical
529 studies and demanding a resolution. It is especially important for archaea, the
530 relatively recently discovered Domain, and which embraces a majority of difficult-
531 to-culture organisms. The cultured diversity of archaea is dramatically low:
532 according to the Euzeby LSPN online resource (http://www.bacterio.net/), only
533 some 116 genera and 451 species with validly published names of archaea (of
534 which 55-60% are haloarchaea-related organisms) *vs* some 2277 genera и 11940
535 species of cultured and described bacteria are known to-date. The acidophiles of
536 the order *Thermoplasmatales* are a good example of this status of things,
537 accounting for only six cultured genera published since 1970, despite numerous
538 documentations on the presence of highly diverse *Thermoplasmatales*-like
539 organisms in low-pH habitats worldwide. The present genomic analysis of new
540 successfully cultured *Thermoplasmatales* members[13] brought us closer to the
541 understanding of functional diversity within this archaeal group. Interestingly, these
542 archaea represent a unique case for *Thermoplasmatales*, when organisms from
543 the same species and almost identical genomes from different geographic
544 locations became cultured. Metabolically, *Cuniculiplasmataceae* resemble other
545 *Thermoplasmatales* members, however certain discrepancies suggest some
546 variety of their evolutionary trajectories. *Cuniculiplasma* spp. genomes encode the
547 A1-type heme-copper oxidases forming a distinct clade at the root of A-type
548 reductases and closely branching to the B-type oxygen reductases and are
549 deficient in membrane-integral oxygen reductase subunits III and IV, suggesting
550 that, in contrast with other *Thermoplasmatales,* they have a more ancient and less

551 energetically efficient B-type enzymes. *Cuniculiplasma* spp. exhibit largest
552 genomes among *Thermoplasmatales* seemingly at the expences of genetic loci for
553 heavy metal resistance and defense systems. Scavenger type of nutrition was
554 confirmed as a characteristic trait for *Cunicuiplasma* spp., which is reflected in their
555 genomic blueprints and physiology, suggesting these organisms feed *in situ* on
556 proteinaceous compounds derived from primary producing organisms. Based on
557 the reconstructions of metagenomic data, the archaea related to this species
558 previously supposed to be uncultured and associated to ’G-plasma cluster‘ are
559 found in many acidic environments[1,6]. Certain features predicted from the
560 metagenomic assembly “G-plasma” have not been confirmed highlighting the
561 essentiality of cultivation efforts and experimental functional validation of genomic
562 predictions. Almost identical genomes of the two European isolates and their North
563 American sibling and strong conservation within their genomic islands, suggest a
564 massive stabilizing selective pressure in similar acidic environments and/or
565 significant fidelity of DNA repair systems assure their genome stability.
566 Isolation of reference strains and experimental validation of genomic predictions for
567 this archaeal group should be considered in the future as tasks of a highest priority.
568

569 **Methods**

570 *Sampling, and culturing, DNA isolation and sequencing*.
571 Samples from acidic streamers for isolation were taken in March-April of 2011 from
572 Cantareras (Spain) and Parys Mountain/Mynydd Parys (UK) copper-containing
573 sulfidic ores. Both cultures were grown in AB Medium, pH 1-1.2, as described
574 previously[13]. DNA was isolated by GNOME DNA Isolation Kit (MP Biomedicals).
575

576 *Genome sequencing and analysis protocol*

577 The genomes were sequenced at Fidelity Systems, Inc. (Gaithersburg, MD) using
578 Illumina HiSeq 2000 platform, combining short paired-end libraries of 400 bp and
579 long mate-paired 3,600 bp inserts with an average read length of 100 bases using
580 manufacturer protocols with the only modification that for the PCR amplification of
581 the genome library the TopoTaq DNA polymerase was used[46]. Initially, Velvet v.
582 1.2.10 was used to assemble the contigs[47]. Scaffolding, filling the gaps, and repeat
583 resolution were performed using the Phred, Phrap, Consed software package[48]
584 and in-house software of Fidelity Systems. The error rate quality score of the
585 completed genome sequences was of Phred 50. The final assemblies provided
586 564- and 561-fold coverages for strain S5 and PM4, correspondingly. The genome
587 annotation was done at Fidelity Systems Ltd. using FgenesB 2.0 (SoftBerry, Inc.,
588 NY) followed by manual curation. The Rfam 11.0 database
589 (http://rfam.sanger.ac.uk)[49] and Infernal 1.0.2 (http://infernal.janelia.org)[50] were
590 used for annotation of RNA genes.

591 For analysis of shared and unique proteins all *in silico* translated genes were
592 filtered by a length of 150 amino acids to exclude false predictions from the
593 analysis. Resulting proteins were subjected to 'all vs all' alignment with blastp
594 algorithm[51] and e-value cut-off of 10^{-5}. Resulting blast table was used as an input
595 for OrthoMCL analysis with grain value of 2.5.

596 Assignment of predicted CDS to the archaeal clusters of orthologous groups
597 (arCOGs) was made using blastp against the latest version of arCOG database[52]
598 with maximal e-value of 10^{-5}. blastp hits were filtered to have minimum alignment
599 length more than 50% of query and subject sequences length. arCOG was
600 assigned to a protein if the hits to at least 3 different genera were registered.

601 Phylogenetic analyses were performed in MEGA 6[53] using Maximum likelihood
602 method and bootstrap confidence test. Sequence alignments were performed in
603 MAFFT v. 7[54].
604 Metagenome data search was performed through the following databases: MG-
605 RAST[55], IMG-M-ER databases[56] and SRA archive[57]. Metagenome sequencing
606 projects related to acidic environment were identified using keywords “acid” “mine”
607 “drainage” “copper” and its combination. *Cuniculiplasma* related sequences were
608 detected using blastn algorithm. Sequences with identity > 95% were considered
609 positive hits for MG-RAST and IMG-M-ER, while for NCBI SRA sequences identity
610 cut-off was set to 99 %. CRISPR repeat sequences were analysed locally using
611 blastn tool of NCBI blast 2.4.0+ package against NCBI nt/nr, env_nt and htgs
612 databases, PM4, S5 and ‘G-plasma’ genomes and against metagenomes acidic
613 environments found in IMG-M database (Gp0051182, Gp0097388, Gp0097859,
614 Gp0097858, Gp0053344 and Gp0053343). Parameters were as follows: word size:
615 7, match score: 1, mismatch penalty: -1, gap open penalty: 10, gap extension
616 penalty: 2, percentage of query covered: 90, percentage identical bases: 90.
617 Genomic islands (GIs) in *C. divulgatum* were inspected using Island Viewer 3[58].
618

619 **Acknowledgements**

620 This work was supported by the Royal Society UK- Russia Exchange Grant #IE
621 130218 (13-04-92606 RFBR) and ERA Net IB2 Project MetaCat through UK
622 Biotechnology and Biological Sciences Research Council (BBSRC) Grant
623 BB/M029085/1. Work of IVK and SNG on metabolic reconstruction of *C. divulgatum*
624 was supported by the Russian Science Foundation (RSF) grant # 14-24-00165.
625

626 **Conflict of interest**

627 The authors declare no conflict of interest.

628

629 **Authors contributions**

630 OVG and PNG convened the research. OVG, IVK, TH, AAK, TYN, SNG, SVT and
631 PNG did genome analysis. OVG, HL, IVK, SNG, SVT and PNG wrote the
632 manuscript.

633

634

635 Supplementary Information is available at the Journal website.

636

637

638 REFERENCES

639 1. Méndez-García, C. *et al.* Microbial diversity and metabolic networks in acid
640 mine drainage habitats. *Front. Microbiol*. **6,** 475 (2015).
641 2. Tyson, G. W. *et al*. Community structure and metabolism through
642 reconstruction of microbial genomes from the environment. *Nature* **428,** 37-
643 43 (2004).
644 3. Chen, L. X. *et al*. Microbial communities, processes and functions in acid mine
645 drainage ecosystems. *Curr. Opin. Biotechnol.* **38,** 150-158 (2016).
646 4. Yelton, A. P. *et al*. Comparative genomics in acid mine drainage biofilm
647 communities reveals metabolic and structural differentiation of co-occurring
648 archaea. *BMC Genomics* **17**, 485 (2013).
649 5. Justice, N. B. *et al*. Heterotrophic archaea contribute to carbon cycling in low-
650 pH, suboxic biofilm communities. *Appl. Environ. Microbiol*. **78,** 8321–8330
651 (2012).
652 6. Baker, B. J. & Banfield, J. F. Microbial communities in acid mine drainage.
653 *FEMS Microbiol. Ecol*. **44**, 139–152 (2003).
654 7. Golyshina, O. V. *et al*. *Ferroplasma acidiphilum* gen. nov., sp. nov., an
655 acidophilic, autotrophic, ferrous-iron-oxidizing, cell-wall-lacking, mesophilic
656 member of the *Ferroplasmaceae* fam. nov., comprising a distinct lineage of
657 the Archaea. *Int. J. Syst. Evol. Microbiol.* **50**, 997-1006 (2000).
658 8. Golyshina, O. V. The Family *Ferroplasmaceae*. *In* The Prokaryotes. (Eds.
659 Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F.) 29-
660 34 (Springer Berlin Heidelberg, 2014).
661 9. Golyshina, O. V. Environmental, biogeographic, and biochemical patterns of
662 archaea of the family *Ferroplasmaceae*. *Appl. Environ. Microbiol.* **77,** 5071-
663 5078 (2011).
664 10. Méndez-García, C. *et al*. Microbial stratification in low pH oxic and suboxic
665 macroscopic growths along an acid mine drainage. *ISME J*. **8,** 1259-1274
666 (2014).
667 11. Mueller, R. et al. Ecological distribution and population physiology defined by
668 proteomics in a natural microbial community. *Mol. Syst. Biol.* **6,** 374 (2010).
669 12. Jones, D. S. *et al.* Community genomic analysis of an extremely acidophilic
670 sulfur-oxidizing biofilm. *ISME J.* **6,** 158-70 (2012).

671 13. Golyshina, O. V. *et al*. The novel, extremely acidophilic, cell wall-deficient
672 archaeon *Cuniculiplasma divulgatum* gen. nov., sp. nov. represents a new
673 Family of *Cuniculiplasmataceae* fam. nov., *order Thermoplasmatales. Int. J.*
674 *Syst. Evol. Microbiol.* **66,** 332-340 (2016).
675 14. Kato, S, Itoh, T. & Yamagishi, A. Archaeal diversity in a terrestrial acidic spring
676 field revealed by a novel PCR primer targeting archaeal 16S rRNA genes.
677 *FEMS Microbiol. Lett*. **319,** 34-43 (2011).
678 15. Bonnefoy, V. & Holmes, D. S. Genomic insights into microbial iron oxidation
679 and iron uptake strategies in extremely acidic environments. *Environ.*
680 *Microbiol*. **14**, 1597-1611 (2012).
681 16. Fütterer, O., *et al*. Genome sequence of *Picrophilus torridus* and its
682 implications for life around pH 0. *Proc. Natl. Acad. Sci. USA* **101,** 9091-9096
683 (2004).
684 17. Golyshina, O. V. *et al*. *Acidiplasma aeolicum* gen. nov., sp. nov., a
685 euryarchaeon of the family *Ferroplasmaceae* isolated from a hydrothermal
686 pool, and transfer of *Ferroplasma cupricumulans* to *Acidiplasma*
687 *cupricumulans* comb. nov. *Int. J. Syst. Evol. Microbiol*. **59**, 2815-2823 (2009).
688 18. Goris, J. *et al*. M. DNA-DNA hybridization values and their relationship to
689 whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**, 81-91
690 (2007).
691 19. Rodriguez-R, L. M. & Konstantinidis, K. T. Bypassing cultivation to identify
692 bacterial species. *ASM Microbe*. **9,** 111-118 (2014).
693 20. Guo, J. *et al.* Horizontal gene transfer in an acid mine drainage microbial
694 community. *BMC Genomics* **16**, 496 (2015).
695 21. Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired
696 virus resistance in natural microbial communities. *Science* **320,** 1047-1050
697 (2008).
698 22. Schneider, G. *et al.* Mobilisation and remobilisation of a large archetypal
699 pathogenicity island of uropathogenic *Escherichia coli in vitro* support the role
700 of conjugation for horizontal transfer of genomic islands. *BMC Microbiol.* **11**,
701 210 (2011).
702 23. Aizenman, E., Engelberg-Kulka H. & Glaser, G. An *Escherichia coli*
703 chromosomal "addiction module" regulated by guanosine corrected 3',5'-

704 bispyrophosphate: a model for programmed bacterial cell death. *Proc. Natl.*
705 *Acad. Sci. USA* **93**, 6059–6063 (1996).
706 24. Christensen, S. K., Mikkelsen, M., Pedersen, K. & Gerdes, K. RelE, a global
707 inhibitor of translation, is activated during nutritional stress. *Proc. Natl. Acad.*
708 *Sci. USA* **98,** 14328–14333 (2001).
709 25. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Comprehensive comparative-
710 genomic analysis of Type 2 toxin-antitoxin systems and related mobile stress
711 response systems in prokaryotes. *Biol. Direct* **4**, 19 (2009).
712 26. Makarova, K. S. *et al*. An updated evolutionary classification of CRISPR-Cas
713 systems. *Nat. Rev. Microbiol.* **11,** 722-736 (2015).
714 27. Dick, G. J., *et al*. Community-wide analysis of microbial genome sequence
715 signatures. *Genome Biol*. **10**, R85 doi: 10.1186/gb-2009-10-8-r85 (2009).
716 28. Pointon, C. R. & Ixer, R. A. Parys Mountain mineral deposit, Anglesey, Wales:
717 geology and ore mineralogy. *Trans. Inst. Mining Metallurgy* (*Section B:*
718 *Applied earth science*) **89,** B143-B155 (1980).
719 29. Tsuda, M., Tan, H. M., Nishi, A. & Furukawa, K. Mobile catabolic genes in
720 bacteria. *J. Biosci. Bioeng.* **87**, 401-410 (1999).
721 30. Desmond, E., Brochier-Armanet, C., & Gribaldo, S. Phylogenomics of the
722 archaeal flagellum: rare horizontal gene transfer in a unique motility structure.
723 *BMC Evol. Biol.* **2**,106 (2007).
724 31. Makarova K. S., Koonin E. V., & Albers S. V. Diversity and evolution of Type
725 IV pili systems in Archaea. *Front. Microbiol.* **7**, 667 (2016).
726 32. Lassak, K., Ghosh, A. & Albers, S. V. Diversity, assembly and regulation of
727 archaeal type IV pili-like and non-type-IV pili-like surface structures. *Res.*
728 *Microbiol*. **163,** 630-644 (2012).
729 33. Barnett, J. P., Eijlander, R. T., Kuipers O. P. & Robinson, C. A. Minimal Tat
730 system from a Gram-positive organism: a bifunctional TatA subunit
731 participates in discrete TatAC and TatA complexes. *J. Biol. Chem.* **283**, 2534–
732 2542 (2008).
733 34. Rawlings, N. D., Waller, M., Barrett, A. J. & Bateman, A. MEROPS: the
734 database of proteolytic enzymes, their substrates and inhibitors. *Nucl. Acids*
735 *Res*. **42,** D503-D509 (2014).

736 35. Lin, X. & Tang, J. Purification, characterization, and gene cloning of
737 thermopsin, a thermostable acid protease from *Sulfolobus acidocaldarius. J.*
738 *Biol. Chem*. **265**, 1490-1495 (1990).
739 36. Saier, M. H., Reddy, V. S., Tsu, B. V., Ahmed, M. S., Li C., & Moreno-
740 Hagelsieb G. The Transporter Classification Database (TCDB): recent
741 advances. *Nucleic Acids. Res.* **44**, D372–9 (2016).
742 37. Baughn, A. D., Garforth, S. J., Vilchèze, C. & Jacobs, W.R. Jr. An anaerobic-
743 type alpha-ketoglutarate ferredoxin oxidoreductase completes the oxidative
744 tricarboxylic acid cycle of *Mycobacterium tuberculosis*. *PLoS Pathog*. **5**,
745 e1000662 (2009).
746 38. Nichols, D. G. & Ferguson, S. J. (eds.) Bioenergetics, 4th edn. (London:
747 Academic Press 2013).
748 39. Borisov, V. B., Gennis, R. B., Hemp, J. & Verkhovsky, M. I. The cytochrome
749 bd respiratory oxygen reductases. *Biochim. Biophys. Acta* **1807**, 1398-1413
750 (2011).
751 40. Sousa, F. L., Alves, R. J., Pereira-Leal, J. B, Teixeira M., & Pereira, M. M. A
752 bioinformatics classifier and database for heme–copper oxygen
753 reductases. *PLoS One* **6,** e19117 (2011).
754 41. Sousa, F. L. *et al*. The superfamily of heme-copper oxygen reductases: types
755 and evolutionary considerations. *Biochim. Biophys. Acta* **1817**, 629-637
756 (2012).
757 42. Ducluzeau, A. L *et al*. W. The evolution of respiratory O_2/NO reductases: an
758 out-of-the-phylogenetic-box perspective. *J. R. Soc. Interface* **6,** 98 (2014).
759 43. Muntyan, M. S. *et al.* Cytochrome *cbb*$_3$ of *Thioalkalivibrio* is a Na^+-pumping
760 cytochrome oxidase. *Proc. Natl. Acad. Sci. USA* **112,** 7695–7700 (2015).
761 44. Kozubal, M. A. Dlakic´, M. Macur, R. E. & Inskeep, W. P. Terminal Oxidase
762 Diversity and Function in “*Metallosphaera yellowstonensis*”: Gene Expression
763 and Protein Modeling Suggest Mechanisms of Fe (II) Oxidation in the
764 *Sulfolobales. Appl. Environ. Microbiol.* **77**, 1844–1853 (2011).
765 45. Castelle, C. J. *et al.* The aerobic respiratory chain of the acidophilic archaeon
766 *Ferroplasma acidiphilum*: A membrane-bound complex oxidizing ferrous iron.
767 *Biochim. Biophys. Acta* **1847**, 717-728 (2015).
768 46. Pavlov, A. R., Pavlova, N. V., Kozyavkin, S. A. & Slesarev, A. I. Cooperation
769 between catalytic and DNA binding domains enhances thermostability and

770 supports DNA synthesis at higher temperatures by thermostable DNA
771 polymerases. *Biochemistry* **51**, 2032–2043 (2012).
772 47. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read
773 assembly using de Bruijn graphs. *Genome Res*. **18,** 821-829 (2008).
774 48. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence
775 finishing. *Genome Res*. **8**, 195–202 (1998).
776 49. Burge, S. W., *et al*. fam. 11.0: 10 years of RNA families. *Nucleic Acids Res*.
777 **41**), D226–D232 (2013).
778 50. Navrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA
779 alignments. *Bioinformatics* **25**, 1335–1337 (2009).
780 51. Altschul, S. F., Gish, W., Miller, W., Myers, E. & Lipman, D. J. Basic local
781 aligment search tool. *J. Mol. Biol*. **215**, 403-410 (1990).
782 52. Wolf, Y. I., Makarova, K. S., Yutin, N. & Koonin, E. V. Updated clusters of
783 orthologous genes for Archaea: a complex ancestor of the Archaea and the
784 byways of horizontal gene transfer. *Biol. Direct* **7,** 46 (2012).
785 53. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6:
786 Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol*. 30,
787 2725-2729 (2013)
788 54. Katoh, K. *et al*. MAFFT: a novel method for rapid multiple sequence
789 alignment based on fast Fourier transform. *Nucleic Acids Res*. **30,** 3059-
790 3066 (2002).
791 55. Meyer, F. *et al*. The metagenomics RAST server - a public resource for the
792 automatic phylogenetic and functional analysis of metagenomes*. BMC*
793 *Bioinformatics.* **9,** 386 (2008).
794 56. Markovitz, V. M. *et al*., IMG 4 version of the integrated microbial genomes
795 comparative analysis system. *Nucleic Acids Res*. **42,** D560–D567 (2014).
796 57. Leinonen, R., Sugawara, H., Shumway, R. & International Nucleotide
797 Sequence Database Collaboration. The Sequence Read Archive. *Nucleic*
798 *Acids Res*. **39,** D19–D21 (2011).
799 58. Dhillon, B.K., *et al.* IslandViewer 3: more flexible, interactive genomic island
800 discovery, visualization and analysis. *Nucleic Acids Res.* **43,** W104-W108
801 (2015).
802

803 **Table 1. Overview of general genomic features of *Cuniculiplasma divulgatum*,**
804 **strains PM4 and S5 and ‘G-plasma’**
805

	Strain PM4	Strain S5	G-plasma*
Number of bases	1878916	1938699	1827255
Number of chromosome contigs	1	1	22
introns	1	5	ND
GC mol %	37.16	37.3	38.9
Coding density, %	87.1	87.4	88.5
Genes	1948	2016	1923
tRNA	46	46	48

806 * “G-plasma” genome stats may be affected by the application of a different annotation
807 pipeline and the fact that it was assembled from metagenomic reads[4] as opposed to the
808 genomic assembly of pure cultures of strains S5 and PM4. ND, not determined.
809
810
811

812 **Figure legends**

813 **Figure 1.** Worldwide distribution of *Cuniculiplasma*-related archaea. Map was
814 created using Plotly online package (https://plot.ly/) using geographical
815 coordinates, retrieved from metadata of database entries.

816

817 **Figure 2.** Electron micrographs of *Cuniculiplasma divulgatum* showing monolayer
818 membranes and absence of the S-layer (a & b), pilus (c, arrow) and pleomorphism
819 of cells. Scale bars: 500 nm (a), 200 nm (b), 1 µm (c, d). Ultrathin sections (a, b)
820 and Pt-C shadow castings (c, d). Figure shows cells of the strain PM4 (b, c and d)
821 and S5 (a). Arrowheads in c and d indicate the direction of shadow cast, arrows in
822 a and b point to the cytoplasmic membrane.

823

824 **Figure 3.** Distribution of arCOG Functional Categories within core, accessory and
825 strain-specific (unique) proteomes of *C. divulgatum* S5 and PM4 and 'G-plasma'.
826 Circle area is proportional to the fraction of corresponding arCOG FC to the total
827 number of arCOG hits in every group of proteins. Core group corresponds to
828 proteins found and all three genomes, accessory group includes proteins found in
829 at least two genomes and unique group consists of strain-specific proteins.

830

831 **Figure 4.** Genomic islands (GIs) in *C. divulgatum* strains PM4 and S5. Rings from
832 outside to inside: genomic coordinates (grey colour); plus-strand CDS (blue) and
833 RNA (red);); minus-strand CDS (blue) and RNA (red); strain-specific CDS (red)
834 and genomis islands (green); blastn hits with e-value cutoff 10^{-5} vs other *C.*
835 *divulgatum* isolate; blastn hits with e-value cutoff 10^{-5} vs 'G-plasma'. Function of

836 GIs is marked by small figures: ‘defense’ islands – squares, ‘transporter’ islands –
837 triangles, islands of non-specific function – circles.
838
839 **Figure 5.** Maximum Likelihood phylogenetic tree of PF00115 polypeptides (COX1
840 family). Totally 112 sequences were used in analysis after 50% sequence identity
841 filtering. The tree with the highest log likelihood (-68482.6023) is shown. The
842 percentages of trees in which the associated taxa clustered together (bootstrap
843 values, 1000 replicates) are shown next to the branches. The tree is drawn to
844 scale, with branch lengths measured in the number of substitutions per site. All
845 positions with less than 95% site coverage were eliminated. Unclassified group
846 was first mentioned[42]. Nitric oxide reductases (NOR) were placed as an outgroup.
847
848 **Figure 6.** Phylogentic position of *Cuniculiplasma* spp. within *Thermoplasmata.*
849 **A.** Maximum Likelihood phylogenetic tree, based on concatenated sequences of 11
850 conservative ribosomal proteins of two *Cuniculiplasma* strains, nine
851 *Thermoplasmata* representatives with the genomes, publically available in IMG and
852 *Methanopyrus kandleri* AV19 as an outgroup (not shown on the tree). The proteins,
853 involved into analysis were: COG0048, ribosomal protein S12; COG0049,
854 ribosomal protein S7; COG0081, ribosomal protein L1; COG0197, ribosomal
855 protein L16/L10AE; COG0200, ribosomal protein L15; COG0244, ribosomal protein
856 L10; COG1631, ribosomal protein L44E; COG1890, ribosomal protein S3AE;
857 COG2004, ribosomal protein S24E; COG2051, ribosomal protein S27E; COG2125,
858 ribosomal protein S6E (S10). The tree with the highest log likelihood (-24738.5123)
859 is shown. The percentages of trees in which the associated taxa clustered together
860 (bootstrap values, 1000 replicates) are shown next at branching points. All

861 positions with less than 95% site coverage were eliminated. There were a total of
862 1607 positions in the final dataset. The tree was constructed in MEGA6[53].
863 **B.** IMG COG-based hierarchical clustering. The analysis was performed using IMG
864 genomic annotations of two *Cuniculiplasma* strains and nine publically available
865 *Thermoplasmata* representatives. Bars indicate the number of substitutions per
866 site.
867