View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UCL Discovery

1 Expression of the arsenite oxidation regulatory operon in *Rhizobium* sp. str. NT-
2 26 is under the control of two promoters that respond to different environmental
3 cues
4
5 Running title: Expression of *aioXSR* controlled by two promoters
6
7 Paula M. Corsini[1], Kenneth T. Walker[1] and Joanne M. Santini[1]
8
9 [1]Institute of Structural and Molecular Biology, Division of Biosciences, University
10 College London, London, WC1E 6BT.
11
12 Corresponding author: j.santini@ucl.ac.uk
13
14
15
16
17
18
19
20
21
22
23
24

25 **Originality-Significance Statement**

26 This is the first study to demonstrate that expression of the *aioXSR* regulatory
27 operon is under the control of two promoters, a RpoD type which allows for
28 constitutive expression and a RpoE2 type that allows for up-regulation of the
29 operon in stationary phase. The latter finding in particular is interesting because
30 *aioXSR* operon gene expression is increased under stressful conditions, when
31 nutrients are limiting, perhaps priming the organism and increasing its ability to
32 detect its food source arsenite in the environment. Interestingly, binding motifs for
33 both the transcriptional regulator AioR and the sigma factor RpoE2 were also
34 identified upstream of genes involved in chemotaxis, quorum sensing and motility
35 further linking these processes with As^{III} oxidation.

36

37 **Summary**

38 *Rhizobium* sp. str. NT-26 is a gram-negative facultative chemolithoautotrophic
39 arsenite oxidiser that has been used as a model organism to study various
40 aspects of arsenite oxidation including the regulation of arsenite oxidation. The
41 three regulatory genes, *aioX*, *aioS* and *aioR*, are co-transcribed when NT-26 was
42 grown in the presence or absence of arsenite. The *aioXSR* operon is up-
43 regulated in stationary phase but not by the presence of arsenite in the growth
44 medium. The two transcription start sites upstream of *aioX* were determined
45 which led to the identification of two promoters, the housekeeping promoter
46 RpoD and the growth-phase dependent promoter RpoE2. Promoter-*lacZ* fusions
47 confirmed their constitutive and stationary phase expressions. The involvement
48 of the NT-26 sigma factor RpoE2 in acting on the NT-26 RpoE2 promoter was

49 confirmed *in vivo* in *E. coli*, which lacks a *rpoE2* homologue, using a strain
50 carrying both the promoter-*lacZ* fusion and the NT-26 *rpoE2* gene. An *in silico*
51 approach was used to search for other RpoE2 promoters and AioR-binding
52 motifs and led to the identification of other genes that could be regulated by
53 these proteins including those involved in quorum sensing, chemotaxis and
54 motility expanding the signalling networks important for the microbial metabolism
55 of arsenite.
56

57 **Introduction**

58 Arsenic (As) is a toxic metalloid and is one of the top 10 chemicals of major
59 public health concern according to the World health organisation (WHO) (WHO,
60 2016). Arsenic in the oxidation states arsenite (As^{III}) and arsenate (As^{V}) are the
61 most common soluble forms found in the environment and both are toxic to
62 organisms (Rosen, 2002). Despite the toxicity of As, a range of phylogenetically
63 diverse prokaryotes are able to survive and thrive in As-contaminated
64 environments (Stolz *et al.*, 2006).

65 As^{III} can serve as an electron donor and is oxidised to the less toxic As^{V}
66 with oxygen as the terminal electron acceptor, anaerobically with nitrate (pH >9)
67 or by anoxygenic photosynthesis (Oremland *et al.*, 2012; Osborne and Santini,
68 2012). Aerobic arsenite oxidation has been observed in many environments and
69 in a phylogenetically diverse range of prokaryotes (Stolz *et al.,* 2006; Osborne
70 and Santini, 2012). In *Rhizobium* sp. str. NT-26 As^{III} can be oxidised
71 autotrophically with carbon dioxide as the sole carbon source or heterotrophically
72 with oxygen as the terminal electron acceptor (Santini *et al.*, 2000).

73 In NT-26, As^{III} is oxidised to As^{V} in the periplasm by the As^{III} oxidase (Aio),
74 which is a bioenergetic enzyme that contains a large catalytic subunit (AioA) with
75 a molybdopterin guanine dinucleotide at its active site and a 3Fe-4S cluster, and
76 a small (AioB) Rieske cluster (Santini and vanden Hoven 2004, Warelow *et al.*,
77 2013). Homologues of the *aioB* and *aioA* genes have been identified in many
78 phylogenetically diverse prokaryotes including members of the Bacteria and
79 Archaea (van Lis *et al.*, 2013). In many cases the *aioB* and *aioA* genes are either
80 upstream or downstream of three regulatory genes, *aioX*, *aioS* and *aioR* (Slyemi
81 *et al.*, 2013). In NT-26, the *aioB* and *aioA* genes are in an operon with *cytC* and
82 *moeA1*, downstream of a RpoN promoter (σ^{54}) and operon expression is induced
83 by As^{III} (Figure 1A) (Santini *et al.*, 2007). The regulatory genes, *aioX*, *aioS* and
84 *aioR* are in a separate operon upstream of *aioB* (Sardiwal *et al.*, 2010), which
85 has been shown to be constitutively expressed (this study). The proposed
86 regulation of *aioB* and *aioA* involves As^{III} sensing by the periplasmic protein AioX
87 and the AioX-As^{III} complex presumably acts as a ligand for the sensor histidine
88 kinase, AioS, which auto-phosphorylates and then phosphorylates the
89 transcriptional regulator, AioR, which binds upstream of the RpoN promoter
90 (TGGCACAACGATTGCA) switching on transcription (Sardiwal *et al.*, 2010; Kang
91 *et al.*, 2012; Liu *et al.*, 2012; Andres *et al.*, 2013).

92 AioR appears to play a wider role in regulating gene expression in arsenite
93 oxidisers. Recently, it has also been shown to positively regulate gene
94 expression of the chemotaxis gene *mcp* in *Agrobacterium tumefaciens* GW4 by
95 binding to the *mcp* regulatory region (Shi *et al.*, 2017). The AioR-binding
96 consensus sequence was also found upstream of the *mcp* gene in NT-26 and

97 *Herminiimonas arsenicoxydans* ULPAs1 (Shi *et al.*, 2017).

98 The overall aim of this work was to get a better understanding of the
99 physiological roles of AioX, AioS and AioR in regulating gene expression in NT-
100 26. To do this, we studied the expression of the *aioX*, *aioS* and *aioR* genes under
101 different growth conditions using Quantitative Reverse Transcription PCR (qRT-
102 PCR). We found that the three genes were co-transcribed and that there was an
103 increase in gene expression in stationary phase. Two transcription start sites
104 were identified upstream of *aioX* which resulted in the discovery of two
105 promoters, RpoD (σ^{70}) and RpoE2 (σ^{24}), that operate under different growth
106 conditions. Promoter functional studies confirmed the differences observed in
107 *aioX, aioS, aioR* gene expression under different growth conditions. *In silico*
108 analyses also implicates the sigma factor RpoE2 in regulating quorum sensing
109 and motility.

110

111 **Results**

112 *Co-transcription of the* aioX*,* aioS *and* aioR *genes in NT-26*

113 To determine whether the *aioX, aioS* and *aioR* genes were co-transcribed in NT-
114 26, and therefore, part of the same operon, RT-PCR was performed using the
115 RNA isolated from NT-26 grown heterotrophically either in the presence or
116 absence of As^{III} (Figure 1B). Two sets of primers were used, one to amplify the
117 3′-end of *aioX* and the 5′-end of *aioS* and, and the second to amplify the 3′-end
118 of *aioS* and the 5′-end of *aioR* (Figure 1A, see arrows). The results demonstrate
119 that *aioX, aioS* and *aioR* are co-transcribed under both conditions.

120 **<u>**Insert Figure 1 here**</u>**

121 *The effect of As^{III} and growth phase on* aioX*,* aioS *or* aioR *expression*

122 As determined by qRT-PCR, As^{III} had no effect on the expression of *aioX*, *aioS*
123 and *aioR* (Figure 2A). The qPCR was normalized using the reference genes that
124 encode glutamine synthetase (*glnA*), citrate synthetase (*gltA*), DNA gyrase
125 subunit B (*gyrB*). These reference genes were selected based on their
126 expression stability in NT-26 when grown heterotrophically in the presence and
127 absence of As^{III} in late-log and stationary phases.

128 In NT-26, the *aioX, aioS* and *aioR genes* were up-regulated in stationary
129 phase of growth when compared to late-log phase. The increase in expression
130 was statistically significant (with $p < 0.05$) with increases of 3.7-fold for *aioX*, 2-
131 fold for *aioS* and 3-fold for *aioR* (Figure 2B and Supplementary file 1).

132 **<u>**Insert Figure 2 here**</u>**

133 *Identification of two transcription start sites and the associated promoters*
134 *upstream of* aioX

135 To determine whether the *aioXSR* operon is under the control of one or more
136 promoters, the transcription start site(s) (TSS) upstream of *aioX* was determined
137 using 5'RACE when NT-26 was grown heterotrophically until late-log or
138 stationary phase. Two different TSS were identified: 1) the proximal one named
139 TSS1 and 2) the distal one named TSS2 (Figure 3A). The untranslated region
140 (UTR) was identified and the TSS corresponds to the first 5′ nucleotide of the
141 UTR. These were named UTR1 and TSS1 for the proximal TSS, highlighted in
142 red and UTR2 and TSS2 for the distal one, highlighted in blue. The presence of

143 two different TSS suggests that this operon is regulated by two different
144 promoters.

145 **Insert Figure 3 about here**

146 To identify the promoters, the two regions upstream of both TSSs were
147 subjected to visual inspection and *in silico* analysis. For the constitutively
148 expressed region, six nucleotides spanning the -10 and -35 regions upstream of
149 TSS1 were selected (Figure 3A) and used to construct the nucleotide pattern
150 [TGGACA-16-TACAGT] (Figure 3B) which has been previously shown to be a
151 consensus sequence for a RpoD promoter (Harley and Reynolds, 1987). This
152 consensus sequence was found upstream of 176 different genes in NT-26
153 (Supplementary file 2) some of which are involved in nitrogen fixation, primary
154 metabolism and other cellular functions associated with RpoD promoters
155 (Ramírez-Romero *et al.*, 2006). Upstream of the putative RpoD promoter we also
156 identified the predicted binding site for AioR (underlined in Figure 3A) (Andres *et*
157 *al.*, 2013).

158 Visual inspection of the region upstream of TSS2 revealed a nucleotide
159 pattern, GGAACN16-17cgTT, similar to the RpoE2-binding site in *Rhizobium*
160 *meliloti* (Figure 3A) (Sauviac *et al.*, 2007). Since the RpoE2-controlled promoters
161 are upregulated during cellular stress (Bastiat *et al.*, 2010), the presence of an
162 RpoE2 promoter binding site might imply that the *aioXSR* operon is upregulated
163 during stationary phase as a general stress response.

164 The putative RpoE2 promoter binding site was used to construct the
165 nucleotide pattern [GGAAC-(N)18-TT-(N)8-G] (Figure 3C) and used to search the
166 NT-26 genome for other genes possibly regulated by the RpoE2 sigma factor.
167 The RpoE2 promoter-binding motif was found upstream of 469 genes in NT-26
168 (Supplementary file 2); such a high number of hits suggests the motif was too
169 generic. Nevertheless, the NT-26 RpoE2 promoter motif was found upstream of
170 the *rpoE2* gene, which is also the case in *R. meliloti* (Sauviac *et al*., 2007). In
171 addition, the RpoE2-binding motif was found upstream of putative genes involved
172 in chemotaxis and motility, *qseB* and *fliG*, respectively.

173

174 *Functional analysis of the RpoD and RpoE2 promoters in NT-26 using reporter*
175 *gene fusion*

176 To verify the function of the RpoD and RpoE2 promoters in regulating the
177 *aioXSR* operon in NT-26, the region upstream of the TSS1 and TSS2,
178 designated P_{aioX1} and P_{aioX2}, were cloned upstream of a promoterless *lacZ* gene
179 in the plasmid pPHU234 (Hübner *et al*., 1991). The plasmid was transferred into
180 NT-26 by conjugation and β-galactosidase activity monitored over the course of
181 growth, with samples taken at early-log (OD_{600} 0.030 – 0.058), mid-log (OD_{600}
182 0.07 – 0.098), late-log (OD_{600} 0.115 – 0.140) and stationary (OD > 0.200) phases
183 (Figure 4).

184 Results of the promoter function assays are reported in Figure 4, with β-
185 galactosidase activity in units plotted against growth phase to observe if there is

186 a link between the growth phase and the activity of the putative promoters. NT-26
187 harbouring the vector pPHU234 alone served as the negative control and as
188 expected there was no detectable β-galactosidase activity (Figure 4). The activity
189 of the P_{aioX2}-*lacZ* gene fusion displayed increased β-galactosidase activity over
190 time with the highest activity detected in stationary phase. As expected the P_{aioX1}-
191 *lacZ* fusion was constitutively expressed with no significant increase in activity
192 over time.

193 **<u>**Insert Figure 4 here**</u>**

194 *The NT-26 RpoE2 is required for activity of the* P_{aiox2} *promoter in Escherichia coli*

195 To confirm the involvement of the NT-26 sigma factor RpoE2 in regulating the
196 *aioXSR* operon, *Escherichia coli,* which doesn't naturally contain a *rpoE2*
197 homologue*,* was used as a host for *in vivo* experiments. *E. coli* containing P_{aioX2}
198 alone showed no detectable *β*-galactosidase activity (Figure 5) however when a
199 plasmid containing the NT-26 *rpoE2* gene was also provided *in trans* β-
200 galactosidase activity was detected (Figure 5).

201 **<u>**Insert Figure 5 here**</u>**

202

203 **Discussion**

204 It has been previously shown that the *aioX*, *aioS* and *aioR* genes are essential
205 for As^{III} oxidation and expression of the arsenite oxidase genes in NT-26
206 (Sardiwal *et al.*, 2010; Andres *et al.*, 2013). The AioX, AioS and AioR proteins are
207 thought to be involved in a three-component system involved in the regulation of

208 the *aioB* and *aioA* genes in the presence of As^{III} in the growth medium (Sardiwal
209 *et al*., 2010; Andres *et al*., 2013). In this study we have shown that the *aioX*, *aioS*
210 and *aioR* genes are co-transcribed and that there is no effect on expression of
211 these genes when NT-26 was grown in the presence of As^{III}. Similar results have
212 also been reported for *Thiomonas arsenitoxydans* 3As, where the presence or
213 absence of As^{III} had no effect on the expression of *aioX, aioS* or *aioR* (Slyemi *et*
214 *al*., 2013). However, in *A. tumefaciens* 5A, the presence of As^{III} induces the
215 expression of *aioX* (Liu *et al*., 2012) and in *H. arsenicoxydans* ULPAS-1, *aioX*,
216 *aioS* and *aioR* are all up-regulated after eight hours exposure to As^{III} (Cleiss-
217 Arnold *et al*., 2010).

218 In NT-26, a growth-phase dependent effect on expression was observed
219 where the *aioX*, *aioS*, and *aioR* genes were up-regulated in stationary phase
220 (Figure 2). These results can be explained by the detection of two different TSSs
221 and the identification of two promoters, RpoD and RpoE2. The TSS upstream of
222 *aioX* in *T. arsenitoxydans* 3As was also determined and a RpoD consensus
223 sequence identified (Moinier *et al*., 2014). Given the constitutive expression of
224 the *aioXSR* operon, the identification of a RpoD promoter was expected. Perhaps
225 surprisingly, the *aioXSR* operon was up-regulated in stationary phase resulting in
226 the identification of a RpoE2 promoter. In *R. meliloti* the RpoE2 sigma factor is
227 involved in the general stress and starvation response (Sauviac *et al*., 2007) and
228 this may also be the case in NT-26.

229 The involvement of RpoE2 in regulating the *aioXSR* operon in stationary
230 phase in NT-26 helps us elucidate the link between the regulation of As^{III}
231 oxidation, motility and quorum sensing that has been previously suggested for

NT-26 (Andres *et al.*, 2013), *Agrobacterium* GW4 (Shi *et al.*, 2017), *A. tumefaciens* 5A (Kashyap *et al.*, 2006) and *H. arsenicoxydans* ULPAs1 (Muller *et al.*, 2007). The RpoE2-binding consensus sequence was also found upstream of the genes *rpoe2*, *kat* and *qseB* (refer to Supplementary file 2), the latter two of which encode putative proteins involved in the response to oxidative stress and flagella regulation, respectively. In NT-26, the *kat* gene was also found to be up-regulated by As^{III} (Andres *et al.*, 2013) and in *S. meliloti* it is known to be regulated by RpoE2 (Sauviac *et al.*, 2007). In *E. coli* The *qseB* gene encodes a putative regulatory protein involved in quorum sensing and flagella gene expression (Sperandio *et al.*, 2002) and was also shown in NT-26 to be up-regulated by As^{III} (Andres *et al.*, 2013). NT-26 was also shown to be more motile when grown in the presence of As^{III}, reinforcing the link presented here between the regulation of *qseB* and the *aioXSR* operon regulation by RpoE2 (Andres *et al.*, 2013). We hypothesise that in stationary phase when the As^{III} concentration is reduced (Santini *et al.*, 2000), that a greater abundance of AioX, AioS and AioR would allow NT-26 to sense and respond to lower concentrations of As^{III}. This together with the As^{III}-induced regulation of chemotaxis would facilitate its mobility towards As^{III}.

In NT-26 the consensus sequences for the predicted AioR-binding site was found upstream of the putative chemoreceptor-encoding gene, *mcp* Shi *et al.*, 2017). Mcp is a chemoreceptor described in *A. tumefaciens* GW4 to bind As^{III} and showed that it was important for chemotaxis towards As^{III} in this organism (Shi *et al.*, 2017). In NT-26 it seems that Mcp may also be involved in chemotaxis towards As^{III} and that *mcp* expression is under the control of AioR, further

256 strengthening the link between As^{III}-sensing, As^{III} oxidation and As^{III}-induced
257 chemotaxis. The chemotaxis genes, involved in regulating the activity and
258 direction of the flagella (Wadhams and Armitage, 2004), *cheY, cheR, cheW* and
259 *cheD* also contain a putative RpoE2-binding site in their promoter regions
260 (Supplementary file 2) (no RpoD promoters were identified).

261 The presence of As^{III} in the growth medium has no effect on the
262 expression of the *aioXSR* operon, however, the results presented herein
263 introduces possible links between the regulation of As^{III} oxidation, motility and
264 quorum-sensing and opens up the possibility that the regulatory proteins AioX,
265 AioS and AioR may have other roles other than regulating As^{III} oxidation in NT-
266 26. The results presented here also suggests the involvement of the sigma factor
267 RpoE2 in the regulation of As^{III} oxidation and the link to chemotaxis and confirms
268 the involvement of RpoD in regulating expression of the *aioXSR* operon as first
269 suggested for *T. arsenitoxydans* 3As (Moinier *et al*., 2014).

270

271 **Experimental Procedures**

272 *Bacterial Culture*

273 A rifampicin resistant (Rif^R) spontaneous mutant of NT-26 (Santini and vanden
274 Hoven, 2004) was grown in McCartney bottles containing 10 ml Minimal Salts
275 Medium (MSM) containing 0.04 % yeast exact (YE) (Oxoid™) with and without 5
276 mM As^{III} (Santini *et al*., 2000). Routine transfers were done using a 5% (v/v)
277 inoculum of NT-26 grown overnight in the respective medium. All cultures were
278 incubated at 28°C under aerobic conditions with shaking at 150 rpm. For the
279 qPCR and promoter activity experiments, the cells were grown until late-log

280 (OD_{600} from 0.100 to 0.140) and stationary phases (OD_{600} from 0.170 to 0.24)
281 (Santini *et al*., 2000). *E. coli* was routinely cultured in lysogeny broth (LB).
282
283 *Nucleic acid isolation*
284 NT-26 genomic DNA (gDNA) was isolated using the Wizard® Genomic DNA
285 purification kit (Promega) according to the manufacturer's instructions.
286 Total RNA was isolated from NT-26 using the SV Total RNA Isolation
287 System (Promega) following the manufacturer's instructions. To avoid DNA
288 contamination, the DNA-*free* Kit™ (Ambion) was used according to the
289 manufacturer's instructions and the RNA stored at -80 °C. RNA was isolated from
290 five biological replicates for each of the conditions tested by qPCR.
291 The plasmid pPHU234 (Hubner *et al*., 1991) and recombinant plasmids
292 were isolated from *E. coli* using the QIAprep Spin Miniprep Kit (QIAGEN)
293 according to the manufacturer's instructions.
294 Nucleic acid concentrations were estimated using a nanodrop
295 spectrophotometer (Thermo Scientific NanoDrop 2000c).
296
297 *RT-PCR*
298 The Access RT-PCR system kit (Promega) was used to confirm the co-
299 transcription of *aioX, aioS* and *aioR* in accordance with the manufacturer's
300 instructions. To confirm that the samples were free of DNA contamination, the RT
301 step was removed and only DNA polymerase was used in the reaction; no PCR
302 products were obtained in these reactions. The primers used in the RT-PCR
303 reactions are listed in the Supplementary file 3.

304 *Real time PCR*

305 qPCR reactions were performed using the PikoReal 96 Real-Time PCR System
306 (Thermo Scientific) using the DyNAmo™ ColorFlash SYBR® Green qPCR Kit
307 (Thermo Scientific). cDNA was synthesised using the RevertAid Premium First
308 Strand cDNA Synthesis kit (Thermo Scientific) and the quantitative PCR (qPCR)
309 First Strand cDNA Synthesis protocol was performed using random primers
310 provided and according to the manufacturer's specifications. The amount of total
311 RNA used to synthesise cDNA was 1 µg and in the qPCR reactions a final
312 concentration of 2 ng/µl of cDNA was used.

313 The baseline and quantification cycle (Cq) of each reaction was
314 automatically determined using PikoReal Software version 2.1 (Thermo
315 Scientific). The software qBase+ (Hellemans *et al*., 2007) was used to analyse
316 the qPCR data to normalise expression levels of the target genes, based on the
317 expression of the reference genes *glnA, gyrB* and *gltA*. qBase+ software was
318 also used to calculate the expression level of each gene in the conditions tested,
319 to plot the results using 95% confidence interval and to perform the analyses of
320 variance (ANOVA). When comparing two different conditions, the expression of a
321 gene was considered significantly different when $p < 0.05$ and not significantly
322 different when $p > 0.05$ based on the ANOVA.

323

324 *Promoter identification*

325 The SMARTer RACE 5′ kit from Clonetech (Takara Bio Inc., Shiga, Japan) was
326 used to determine the TSS upstream of *aioX* following the manufacturer's
327 specifications. All the reagents, cells, enzymes and vectors used were provided

328 in the kit apart from the gene specific primers designed for *aioX* (Supplementary
329 file 3).

330 DNA samples were sequenced by GATC Biotech (Germany) using the
331 LIGHTrun^{TM} Sanger technology (GATC Biotech AG). MEGA 6.0 (Tamura *et al*.,
332 2013) was used to analyse the sequence chromatograms.

333 To identify the TSS, sequences obtained from the different cloned
334 fragments (four for each TSS) were aligned to the NT-26 *aioX* sequence
335 obtained in the MAGE interface (ID: NT26v4_p10026) (Vallenet *et al*., 2006)
336 using ClustalW (Thompson *et al*., 1994).

337

338 *Promoter activity*

339 To test whether the putative promoters upstream of *aioX* were functional*,* the
340 P_{aioX1} and P_{aioX2} fragments were PCR amplified using the primers *aioX* p1 and
341 *aioX* p2, forward and reverse, containing BamHI and PstI restriction sites and
342 cloned upstream of a promoterless *lacZ* gene in the plasmid pPHU234 at the
343 BamHI/PstI sites (Hübner *et al*., 1991). The plasmids were transferred into NT-26
344 Rif^{R} by conjugation as described previously (Santini and vanden Hoven, 2004).
345 The promoter assays were performed in *E. coli* strain S17 λ *pir*. To quantify the
346 promoter function, β-galactosidase activity was measured as described
347 previously (Zhang and Bremer, 1995) and this was done with three biological
348 replicates (Supplementary file 4).

349

350

351 In silico *searches for RpoD- and RpoE2-regulated genes*

352 For the region identified as P_{aioX1}, the sequences at -35 and -10 were used to
353 build the pattern [TGGACA-(N)16-TACAGT]. For the region identified as P_{aioX2},
354 the RpoE2-binding motif described previously for *S. meliloti* (Sauviac *et al*., 2007)
355 was identified by eye and used to construct the pattern [GGAAC-(N)18-TT-(N)8-
356 G]. The two motifs were used as input for the 'Find Individual Motif Occurrences'
357 tool (FIMO) (Grant *et al*., 2011), which matched the motifs against a library of
358 upstream regions (up to 400 nucleotides in length from the starting ATG) for
359 every gene in NT-26. The resulting alignments were used to generate a summary
360 motif with MEME/MAST (Bailey and Elkan, 1994).

361

362 **Acknowledgments**

363 The research was partly supported by CNPq (Conselho Nacional de
364 Desenvolvimento Científico e Tecnológico) Foundation (Brazil). The authors
365 declare no conflict of interest.

366

367 **References**

368 Andres, J., Arsène-Ploetze, F., Barbe, V., Brochier-Armanet, C., Cleiss-Arnold,
369 J., Coppée, J. Y., *et al*. (2013) Life in an arsenic-containing gold mine: Genome
370 and physiology of the autotrophic arsenite-oxidizing bacterium *Rhizobium* sp. NT-
371 26. Genome Biol Evol 5: 934–953.

372

373 Bailey, T. L., and Elkan, C. (1994) Fitting a mixture model by expectation
374 maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol

375 2: 28–36.

376

377 Bastiat, B., Sauviac, L., and Bruand, C. (2010) Dual control of *Sinorhizobium*

378 *meliloti* RpoE2 sigma factor activity by two PhyR-type two-component response

379 regulators. J. Bacteriol 192: 2255–2265.

380

381 Cleiss-Arnold, J., Koechler, S., Proux, C., Fardeau, M.-L., Dillies, M.-A., Coppee,

382 J.-Y., *et al*. (2010) Temporal transcriptomic response during arsenic stress in

383 *Herminiimonas arsenicoxydans*. BMC Genomics 11: 709.

384

385 Grant, C. E., Bailey, T. L., and Noble, W. S. (2011) FIMO: scanning for

386 occurrences of a given motif. Bioinformatics 27: 1017–1018.

387

388 Harley, C. B., and Reynolds, R. P. (1987) Analysis of *E. coli* promoter

389 sequences. Nucleic Acids Res 15: 2343–2361.

390

391 Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J.

392 (2007) qBase relative quantification framework and software for management

393 and automated analysis of real-time quantitative PCR data. Genome Biol 8: R19.

394

395 Hubner, P., Willison, J. C., Vignais, P. M., and Bickle, T. A. (1991) Expression of

396 regulatory nif genes in *Rhodobacter capsulatus*. J. Bacteriol173: 2993–2999.

397

398 Kang, Y. S., Bothner, B., Rensing, C., and McDermott, T. R. (2012) Involvement

399 of RpoN in regulating bacterial arsenite oxidation. Appl Environ Microbiol 78:
400 5638–5645.
401
402 Kashyap, D. R., Botero, L. M., Franck, W. L., Hassett, D. J., and McDermott, T.
403 R. (2006) Complex regulation of arsenite oxidation in *Agrobacterium*
404 *tumefaciens*. J. Bacteriol 188: 1081–1088.
405
406 Liu, G., Liu, M., Kim, E. H., Maaty, W. S., Bothner, B., Lei, B., *et al.* (2012) A
407 periplasmic arsenite-binding protein involved in regulating arsenite oxidation.
408 Environ Microbiol 14: 1624–1634.
409
410 Moinier, D., Slyemi, D., Byrne, D., Lignon, S., Lebrun, R., Talla, E., and
411 Bonnefoy, V. (2014) An ArsR/SmtB family member is involved in the regulation
412 by arsenic of the arsenite oxidase operon in *Thiomonas arsenitoxydans*. Appl
413 Environ Microbiol 80: 6413-6426.
414
415 Muller, D., Médigue, C., Koechler, S., Barbe, V., Barakat, M., Talla, E., *et al.*
416 (2007) A tale of two oxidation states: Bacterial colonization of arsenic-rich
417 environments. PLoS Genetics 3: 0518–0530
418
419 Oremland, R. ., Stolz, J. F., and Saltikov, C. W. (2012) Anaerobic oxidation of
420 arsenite by autotrophic bacteria: The view from Mono Lake, California, in The
421 Metabolism of Arsenite. Santini, J. M. and Ward, S. A. (eds). Leiden, The
422 Netherlands: CRC Press/Balkema, pp. 73–80.

423 Osborne, T. H., and Santini, J. M. (2012) Prokaryotic aerobic oxidation of
424 arsenite, in The Metabolism of Arsenite. Santini, J. M. and Ward, S. A. (eds).
425 Leiden, The Netherlands: CRC Press/Balkema, pp. 61–72.
426
427 Ramírez-Romero, M. A., Masulis, I., Cevallos, M. A., González, V., and Dávila,
428 G. (2006) The *Rhizobium etli* σ70 (SigA) factor recognizes a lax consensus
429 promoter. Nucleic Acids Res 34: 1470–1480
430 Rosen, B. P. (2002) Biochemistry of arsenic detoxification. *FEBS Letters* 529:
431 86–92.
432
433 Santini, J. M., and vanden Hoven, R. N. (2004) Molybdenum-Containing Arsenite
434 Oxidase of the Chemolithoautotrophic Arsenite Oxidizer NT-26. J. Bacteriol 186:
435 1614–1619.
436
437 Santini, J. M., Kappler, U., Ward, S. A., Honeychurch, M. J., vanden Hoven, R.
438 N., and Bernhardt, P. V. (2007) The NT-26 cytochrome c552 and its role in
439 arsenite oxidation. Biochim Biophys Acta - Bioenergetics 1767: 189–196.
440
441 Santini, J. M., Sly, L. I., Schnagl, R. D., and Macy, J. M. (2000) A new
442 chemolithoautotrophic arsenite-oxidizing bacterium isolated from a gold mine:
443 Phylogenetic, physiological, and preliminary biochemical studies. Appl
444 Environ Microbiol 66: 92–97.
445
446 Sardiwal, S., Santini, J. M., Osborne, T. H., and Djordjevic, S. (2010)

447 Characterization of a two-component signal transduction system that controls
448 arsenite oxidation in the chemolithoautotroph NT-26. FEMS Microbiol Lett 313:
449 20–28.
450
451 Sauviac, L., Philippe, H., Phok, K., and Bruand, C. (2007) An extracytoplasmic
452 function sigma factor acts as a general stress response regulator in
453 *Sinorhizobium melilot*i. J. Bacteriol 189: 4204–4216.
454
455 Shi, K., Fan, X., Qiao, Z., Han, Y., McDermott, T. R., Wang, Q., and Wang, G.
456 (2017) Arsenite oxidation regulator AioR regulates bacterial chemotaxis towards
457 arsenite in *Agrobacterium tumefaciens* GW4. Sci Rep 7: 43252.
458
459 Slyemi, D., Moinier, D., Talla, E., and Bonnefoy, V. (2013) Organization and
460 regulation of the arsenite oxidase operon of the moderately acidophilic and
461 facultative chemoautotrophic *Thiomonas arsenitoxydans*. Extremophiles 17:
462 911–920.
463
464 Sperandio, V., Torres, A. G., and Kaper, J. B. (2002) Quorum sensing
465 *Escherichia coli* regulators B and C (QseBC): A novel two-component regulatory
466 system involved in the regulation of flagella and motility by quorum sensing in *E.*
467 *coli*. Mol Microbiol 43: 809–821.
468
469 Stolz, J. F., Basu, P., Santini, J. M., and Oremland, R. S. (2006) Arsenic and
470 selenium in microbial metabolism. Annu Rev Microbiol 60: 107–30.

471 Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6:
472 Molecular evolutionary genetics analysis version 6.0. Mol Bio Evol 30: 2725–
473 2729.
474
475 Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W:
476 improving the sensitivity of progressive multiple sequence alignment through
477 sequence weighting, position-specific gap penalties and weight matrix choice.
478 Nucleic Acids Res 22: 4673–4680.
479
480 Vallenet, D., Labarre, L., Rouy, Z., Barbe, V., Bocs, S., Cruveiller, S., *et al.*
481 (2006) MaGe: A microbial genome annotation system supported by synteny
482 results. Nucleic Acids Res, 34: 53–65.
483
484 Wadhams, G. H., and Armitage, J. P. (2004) Making sense of it all: bacterial
485 chemotaxis. Nat Rev Mol Cell Biol 5: 1024–1037.
486
487 Warelow, T. P., Oke, M., Schoepp-Cothenet, B., Dahl, J. U., Bruselat, N.,
488 Sivalingam, G. N., *et al.* (2013) The Respiratory Arsenite Oxidase: Structure and
489 the Role of Residues Surrounding the Rieske Cluster. PLoS ONE 8: e72535
490
491 WHO (2016) *Arsenic*. Available at:
492 http://www.who.int/mediacentre/factsheets/fs372/en/
493
494 Zhang, X., and Bremer, H. (1995) Control of the *Escherichia coli* rrnB P1

495 promoter strength by ppGpp. J Biol Chem 270: 11181–11189.

496

497 **Figures legends**

498 **Fig. 1**. Organisation of the NT-26 *aio* gene cluster. **A)** *aioX* encodes periplasmic
499 As^{III}-binding protein; *aioS* encodes sensor histidine kinase; *aioR* encodes
500 transcriptional regulator; *aioB* encodes small subunit of the As^{III} oxidase; *aioA*
501 encodes the large catalytic subunit of the As^{III} oxidase; *cytC* encodes a
502 cytochrome *c*; *moeA*1 encodes a molybdenum cofactor biosynthesis protein. **B)**
503 RT-PCR analysis of the co-transcription of *aioX* - *aioS*, and *aioS*- *aioR* using
504 *aioXF-aioSR* and *aioSF-aioRR* primers (Supplementary file 3) and RNA isolated
505 from NT-26 grown to late log phase with and without As^{III}.

506

507 **Fig. 2.** Relative expression analysis using qPCR to compare the expression
508 ratios **A)** Samples from NT-26 grown heterotrophically (with 0.04% yeast extact,
509 YE) with and without As^{III} in the growth medium. The error bars show the 95%
510 upper and lower confidence intervals and *p> 0.05; **B)** Samples from NT-26
511 grown heterotrophically without As^{III} grown to late-log or stationary phase. The
512 error bars show the 95% upper and lower confidence intervals. **p< 0.05.

513 **Fig. 3.** Identified promoters upstream of *aioX*. **A)** The two identified TSSs
514 upstream of *aioX* are underlined, the TSS1 is underlined in red and the TSS2 is
515 underlined in blue. The six nucleotides flanking the -10 and -35 regions upstream
516 of TSS1 correspond to the RpoD conserved region are in red and in bold. The
517 predicted AioR-binding site is shown italicised and underlined. The conserved
518 regions for the RpoE2 promoter are shown in blue and in bold; **B)** The RpoD
519 promoter conserved site motif built with MEME/MAST using conserved
520 nucleotides in NT-26; **C)** The RpoE2 promoter conserved site motif built with
521 MEME/MAST using conserved nucleotides in NT-26.

522 **Fig. 4.** β-galactosidase activity determined at different growth stages for the
523 *lacZ*–promoter in NT-26 grown heterotrophically (with 0.04% yeast extract, YE)

524 with and without As^{III}. The negative control samples (i.e., the promoterless
525 pPHU234 plasmid) are shown in black. The β-galactosidase activity for NT-26
526 containing the RpoE2 promoter (P_{aioX1}) *in trans* is shown in blue and NT-26
527 containing the RpoD promoter (P_{aioX2}) *in trans* is shown in red. The data plotted
528 correspond to the average of three independent experiments.

529 **Fig. 5.** β-galactosidase activity determined for the *lacZ*–promoter fusions also
530 harbouring the NT-26 *rpoE2* gene in *E. coli*. C, negative control corresponds to
531 the activity of the promoterless plasmid pPHU234. P*aioX1*, negative control of
532 the RpoE2 promoter alone. P_{aioX1} + RpoE2, RpoE2 promoter and also a plasmid
533 harbouring the NT-26 *rpoE2* gene. The *β*-galactosidase activity plotted is the
534 average of 3 independent experiments.

535

536 **Additional files list**

537 File name: Supplementary file 1
538 Description: Raw data for qPCR
539
540 File: Supplementary file 2
541 Description: Results from promoter prediction in NT-26's chromosome and
542 plasmid 1
543
544 File: Supplementary file 3
545 Description: List of primers and strains used in this study
546
547 File: Supplementary file 4
548 Description: Raw data for β-galactosidase activity in NT-26 and *E. coli*
549