CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Apollo
Copyright © 2016 Rosconi et al.
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Accepted Manuscript Posted Online

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

1 **Essential genes for *in vitro* growth of the endophyte *Herbaspirillum***
2 ***seropedicae* SmR1 revealed by transposon insertion sequencing**
3
4 **Running title:** Tn-seq applied to *Herbaspirillum seropedicae*
5
6 Federico Rosconi[1,2a], Stefan P.W. de Vries[2], Abiyad Baig[2b], Elena Fabiano[1],
7 Andrew J. Grant[2*].
8
9 [1]Departamento de Bioquímica y Genómica Microbianas, IIBCE, Montevideo,
10 Uruguay.
11 [2]Department of Veterinary Medicine, University of Cambridge, Cambridge, United
12 Kingdom.
13 [a]Current address: Department of Biology, Boston College, Boston, United States.
14 [b]Current address: School of Veterinary Medicine and Science, University of
15 Nottingham, Sutton Bonnington, Leicestershire, United Kingdom.
16
17 [*]To whom correspondence should be addressed: ajg60@cam.ac.uk.
18
19 Keywords: *Herbaspirillum*, endophytes, Tn-seq.

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

20 **ABSTRACT**

21 **The interior of plants contains microorganisms (referred to as endophytes)**
22 **distinct to those present in the root surface or in the surrounding soil.**
23 ***Herbaspirillum seropedicae* strain SmR1, belonging to the β-**
24 **Proteobacterium, is an endophyte that colonizes crops, including rice, maize,**
25 **sugarcane and sorghum. Different approaches have revealed genes and**
26 **pathways regulated during the interaction of *H. seropedicae* with its plant**
27 **hosts. However, functional genomic analysis of transposon (Tn) mutants has**
28 **been hampered by the lack of genetic tools. Here we successfully employed**
29 **a combination of *in vivo* high-density *mariner* Tn mutagenesis and targeted**
30 **sequencing of Tn insertion sites (Tn-seq) in *H. seropedicae* SmR1. The**
31 **analysis of multiple gene-saturating Tn libraries revealed that 395 genes are**
32 **essential for growth of *H. seropedicae* SmR1 in TY medium. A comparative**
33 **analysis with the "database of essential genes" known as DEG, showed that**
34 **25 genes are uniquely essential in *H. seropedicae* SmR1. The developed Tn**
35 **mutagenesis protocol and the generated gene-saturating Tn libraries will**
36 **assist with the elucidation of the genetic mechanisms of *H. seropedicae***
37 **endophytic lifestyle.**

38

39 **IMPORTANCE**

40 **A focal point in the study of endophytes is the development of effective**
41 **biofertilizers that could help to reduce the input of agrochemicals in**
42 **croplands. Besides plant growth promotion abilities, a good biofertilizer**
43 **should be successful in colonizing its host and compete against the native**

Accepted Manuscript Posted Online

44 **microbiota. By using a systematic transposon (Tn)-based gene-inactivation**
45 **strategy and massively parallel sequencing of Tn insertion sites (Tn-seq) it is**
46 **possible to study the fitness of thousands of Tn mutants in a single**
47 **experiment. We have applied the combination of these techniques to the**
48 **plant growth promoting endophyte *Herbaspirillum seropedicae* strain SmR1.**
49 **The Tn mutant libraries generated will enable studies into the genetic**
50 **mechanisms of *H. seropedicae*-plant interactions. The approach that we have**
51 **taken is applicable to other plant-interacting bacteria.**

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

52 **INTRODUCTION**

53 Plants rely on beneficial interactions with their microbiota for nutrient availability,
54 growth promotion and suppression of disease. The plant interior, referred to as the
55 endosphere has been shown to contain a distinct microbiome, that is less diverse
56 than those from the rhizoplane (root surface) and the rhizosphere (narrow zone of
57 soil subject to the influence of living roots) (1). Microorganisms that colonise the
58 endosphere are referred to as endophytes (2, 3), this includes all microorganisms
59 that for all or part of their lifetime colonise internal plant tissues (4).

60 The knowledge of plant-bacterial endophyte interactions at the genetic and
61 molecular level has increased due to the use of suitable (laboratory controlled)
62 biological models. A model endophyte is *Herbaspirillum seropedicae*, a member of
63 the β-Proteobacteria subclass that includes many plant-associated bacteria such
64 as species of the genera *Azoarcus*, *Burkholderia* and *Ralstonia* (5). Several
65 characteristics make *H. seropedicae* a suitable model endophyte (6); (*i*) it provides
66 fixed nitrogen for important agro-economic cultivars; (*ii*) it is genetically tractable;
67 (*iii*) it has mechanisms of plant growth promotion other than nitrogen fixation; (*iv*) it
68 has a wide range of plant-hosts; (*v*) culturable-bacteria are not isolated from soil,
69 only from the inside of plants (7, 8); and (*vi*) there are publicly-available genome
70 sequences (8). Some isolates of *H. seropedicae* have been described as
71 pathogenic in plants, although this may be the result of the host being unable to
72 control colonisation, there have also been reports that it can be an opportunistic
73 pathogen in immunocompromised individuals (9, 10). The most well-studied *H.*
74 *seropedicae* strains, SmR1 and Z67, have been tested in different plant species
75 without symptoms of disease (11).

Accepted Manuscript Posted Online
Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology

Recently, transcriptomic and proteomic approaches have identified genes and pathways that are regulated during the interaction of *H. seropedicae* with different plant hosts (12-14). In addition, comparative genomics and metagenomics studies have shown that certain functions are overrepresented in bacterial endophytes genomes compared to rhizospheric or soil bacteria, *e.g.* nutrient transport systems, type IV conjugal DNA-protein transfer secretion systems, plant growth promotion genes, and iron uptake systems (4, 15-17). Gene inactivation/deletion studies have shown that lipopolysaccharide (LPS) production is essential for effective *H. seropedicae* attachment to maize roots (18), and high affinity iron-uptake mechanisms contribute to competitive fitness of *H. seropedicae* inside host plants (19).

Compared to gene expression and comparative genomics studies, high-throughput functional analyses of endophyte-plant interactions have lagged behind. In recent years, there has been much progress in the application of transposon-based gene-inactivation methods in combination with massively parallel sequencing of transposon (Tn) insertion sites, *e.g.*, Tn-seq and related techniques (20-22), which have, and continue to, advanced the characterisation of bacterial-host interactions.

In this study we successfully employed *in vivo mariner* Tn mutagenesis in *H. seropedicae* strain SmR1 and characterised the resulting Tn mutants by Tn-seq. The resulting dataset was used to identify the genes that upon inactivation have a detrimental effect on fitness during *in vitro* growth and survival, referred to as essential genes.

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

99 **MATERIALS AND METHODS**

100 **Bacterial strains, media and growth conditions.** *H. seropedicae* SmR1 was
101 routinely cultured at 30°C in TY media (tryptone 5 g l^{-1}, yeast extract 3 gl^{-1}, $CaCl_2$
102 0.1 gl^{-1}). *Escherichia coli* strains NEB 5-alpha (New England Biolabs),
103 TransforMax™ EC100D™ *pir*$^+$ (Epicentre) and SM10-λpir were cultured at 37°C in
104 Luria-Bertaini (LB) broth. Where necessary for selection of plasmids, media was
105 supplemented with ampicillin (50 μg/ml), kanamycin (50 μg/ml) or tetracycline (10
106 μg/ml) as appropriate. For selection of *H. seropedicae* SmR1 Tn mutants, TY
107 media was supplemented with streptomycin (100 μg/ml), and either kanamycin
108 (200 μg/ml) or tetracycline (10 μg/ml) as appropriate. (Bacterial strains and
109 plasmids used in this study are listed in Table 1).

110 **Recombinant DNA techniques.** Standard methods were used for
111 molecular cloning (23). Chromosomal and plasmid DNA purification, DNA
112 modification and ligations were performed using commercial kits according to the
113 manufacturers' instructions (QIAGEN, Thermo Scientific, New England Biolabs).
114 DNA concentration was measured using a Nanodrop ND-1000 spectrophotometer
115 (Thermo Scientific). PCR primers were purchased from Sigma (Sigma-Genosys).
116 Thermal cycling was performed in a Gene Amp® PCR System 9700 (PE Applied
117 Biosystems) or T100™ Thermal Cycler (Bio-Rad). Thermal cycling conditions were
118 96°C for 2 min, then 30 cycles at 96°C for 1 min, 55-60°C for 1 min and 72°C for
119 30 sec/kb, and finally an extension at 72°C for 5 min.

120 **Generation of Tn mutant libraries.** For construction of Tn mutants we
121 used either (*i*) plasmid pSAM_R1 (24), that contains the *mariner* Tn with the

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

122 kanamycin resistance gene *nptII* and the *Himar1_C9* transposase gene under the
123 control of the *rpoD* promoter of the α-Proteobacterium *Rhizobium leguminosarum*,
124 or (*ii*) plasmid pSAM_R5, where we replaced the *nptII* gene with the *tet* gene of
125 plasmid pBR322 flanked by *mariner* specific inverted repeats. The *tetA* gene from
126 pBR322 was amplified with primers Tet_FW1_XhoI and Tet_RV1_XbaI (details of
127 oligonucleotides used in this study are in Table 2). PCR amplicons were cloned
128 into pMiniT using the NEB PCR Cloning Kit (New England Biolabs) generating
129 plasmid pFRC002. This plasmid was digested with *Xho*I and *Xba*I (New England
130 Biolabs), releasing a fragment that was gel purified and cloned into the same
131 restriction enzyme sites of pSAM_R1, generating pSAM_R5. The sequence of
132 these plasmids was confirmed by Sanger sequencing (Source BioScience
133 LifeSciences). Subsequently, the plasmids were transformed into *E. coli*
134 TransforMax™ EC100D™ *pir*$^+$ (Epicentre).

135 Tn mutagenesis was performed by biparental mating using *E. coli* SM10-λpir
136 containing pSAM_R1 or pSAM_R5 as a donor strain as previously described (25).
137 Briefly, 10 mL of a *H. seropedicae* culture was mixed with 5 mL of *E. coli* SM10-
138 λpir (containing pSAM_R1 or pSAM_R5) both at an OD_{600nm} of 0.8. Bacterial cells
139 were washed once with phosphate buffered saline (PBS) (Sigma) and
140 resuspended in 1.5 mL of PBS. 100 µl of this suspension was spotted on TY plates
141 without antibiotics, left to dry and incubated at 30°C overnight. Bacterial colonies
142 were scrapped from the plates and pooled in 10 mL of TY. 100 µl from this
143 suspension was plated on TY agar with streptomycin and depending on the
144 resistance cassette on the Tn element used, either kanamycin or tetracycline was
145 added, and the plates were incubated at 30°C overnight. Bacterial colonies were

Accepted Manuscript Posted Online
Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology
Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

146 scrapped from the plates into 2 mL of TY per plate and pooled. The cell
147 suspension was diluted in 50 mL of TY with the adequate antibiotics at an initial
148 density of 1.5 x 10^7 cells per mL (OD_{600nm} 0.15) and grown shaking to an OD_{600nm}
149 of 0.5. Aliquots were mixed with glycerol to a final concentration of 15% and then
150 frozen at -80°C for future use.

151 As a quality control, the random insertion of Tn's into the chromosome was
152 analysed using a single primer PCR amplification approach to map the Tn insertion
153 site, described in (27). Briefly, randomly selected colonies were isolated from the
154 first Tn library, genomic DNA was extracted and the Tn insertion site was amplified
155 by PCR using the primer Km_RV2 (Table 2) followed by Sanger DNA sequencing
156 of the amplicon. The sequence was aligned to the *H. seropedicae* SmR1 genome
157 to identify the Tn insertion site.

158 **Characterisation of Tn mutant libraries by Tn-seq.** Tn mutant libraries
159 were characterised by Tn-seq, essentially as described previously (28, 29). Briefly,
160 genomic DNA from Tn mutant libraries was isolated using the DNeasy Blood &
161 Tissue Kit (Qiagen), but cells were washed once with 1M NaCl and once in PBS
162 before following the manufacturer's recommendations for Gram-negative bacteria.
163 Five micrograms of genomic DNA was digested with the restriction enzyme *Mme*I,
164 and double-stranded Tn-seq DNA adapters with different barcodes were ligated to
165 the restriction fragments. Tn insertion site flanking sequences were amplified by
166 PCR using adapter- and *mariner* Tn-specific primers using NEBNext Q5 High-
167 Fidelity DNA polymerase (New England Biolabs). Clean-up of the PCR products
168 was performed using MinElute PCR purification columns (Qiagen), DNA

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

169 concentration was measured by Qubit (Life Technologies) and sequenced using 50
170 bp single end sequencing on a HiSeq 2500 Illumina sequencing platform
171 (**Genomics core facility at Cancer Research UK).**

172 **Sequencing data.** Illumina Tn-seq sequencing date has been deposited in
173 the European Nucleotide Archive (http://www.ebi.ac.uk/ena) and are available *via*
174 the study accession number (ENA submission is in process)

175 **Identification of genes essential for *in vitro* growth and survival.** Tn-seq
176 Illumina sequence reads were demultiplexed using the FastX toolkit barcode
177 splitter and analysed using the ESSENTIALS pipeline (30). The following analysis
178 parametere were used in the ESSENTIALS analysis: sequence reads were aligned
179 with a minimal match of 16 nt, repeat regions were filtered, reads mapping to the 3'
180 end of the gene were removed, genomic position bias was corrected through
181 Loess normalisation, read counts were normalised with the trimmed mean of M-
182 values normalisation method (TMM). In the implemented EdgeR statistical analysis
183 part of ESSENTIALS, the dispersion was estimated with the Cox-Reid profile-
184 adjusted likelihood method and the variance was modelled using common
185 dispersion. To determine the number of unique Tn insertion mutants in each
186 library, a read count cut-off was derived from Kernel density plots in R, which allow
187 delineation of "true" Tn insertions from "noise" sequencing reads. The distribution
188 of Tn insertions was visualised by plotting the $\log_2$ read count per chromosomal
189 position using an in-house perl script. As a measure of gene essentiality, the $\log_2$
190 fold-change between the observed and expected sequence reads was calculated
191 per gene and a cut-off was determined as described in (30). Genes that had no

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

192 informative TA insertion site flanking sequences (43 genes), *i.e.* no unique flanking
193 sequence, were excluded from the analysis; for reference these genes are listed in
194 Table S1. Additional selection criteria were as follows: an Benjamini & Hochberg
195 adjusted $P < 0.05$ and a probability that the gene was hit by a Tn of > 0.95, as
196 calculated using a derivative of Poisson's law; $1\text{-}e^{N \times \ln(1-f)}$, with N being the number
197 of unique Tn insertion mutants and f representing the number of unique TA flanking
198 sequences in a gene divided by the number of unique TA flanking sequences in
199 the genome. In addition, genes for which no sequence reads were detected and
200 the probability of disruption was > 0.95 were also considered for further analysis.
201 Functional class enrichment of candidate genes was performed using a Fisher-
202 exact test and corrected for multiple testing using Q-value (31). Genes required for
203 *in vitro* growth and survival of *H. Seropedicae* were visualised in DNAplotter (32).

204 **Essential genes features description.** A homology search of the
205 Database of Essential Genes (DEG) was performed using the BlastP tool at
206 www.essentialgene.org. Other analyses (COG category assignment, metabolic
207 pathways description, and prediction of transmembrane domains and signal
208 peptides) were performed with tools of the Integrated Microbial Genomics platform
209 at http://img.jgi.doe.gov/ (33).

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

RESULTS AND DISCUSSION

Characterisation of *H. seropedicae* SmR1 Tn mutant libraries. To identify genes critical for growth of *H. seropedicae,* Tn mutant libraries were constructed under nutrient-rich conditions, *i.e.* in TY medium, using a biparental mating protocol. The *in vivo* Tn mutagenesis had an efficiency of ~5 x 10^{-6} Tn mutants per *H. seropedicae* recipient cell. A total of six Tn mutant libraries were constructed with sizes ranging between 24,000 and 140,000 colony forming units (CFU) (Table 3).

Tn insertion site sequencing (Tn-seq) was performed using Illumina sequencing. Out of the 88,320 potential TA dinucleotide *mariner* Tn insertion sites in the *H. seropedicae* SmR1 genome, 56,174 insertion sites (*i.e.* 63.6% of the total TA sites) were hit by a Tn insertion (Table 3). A cumulative analysis of amalgamating libraries revealed that the number of new unique Tn insertion mutants leveled off at ~55,000 mutants (Fig. 1A). This suggests that although we only achieved Tn insertions in ~64% of the potential TA dinucleotide *mariner* Tn insertion sites, the maximum empirical number of mutants was obtained by this approach (without the use of much larger libraries). In addition, rarefaction analysis showed that we reached saturation in terms of the number of genes that could be mutated in the *H. seropedicae* genome (Fig. 1B). Tn insertions were distributed evenly throughout the chromosome, without any apparent evidence of hot-spotting, with an average of one Tn insertion every 95 bp (Fig. 1C).

It is widely assumed that genes with very few, or no, Tn insertions represent genes that are essential for growth and survival or were underrepresented because their corresponding Tn insertion mutants have a growth defect (20), or they were

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

234 not inactivated by a Tn element during the Tn mutagenesis. To identify the genes
235 required for growth under nutrient-rich conditions, a fold-change was calculated
236 between the actual number of sequence reads and the number of expected
237 sequence reads (Fig. 2A), the latter considers the number of Tn mutants in the
238 library, the length of the gene and the number of possible Tn insertion positions
239 (*i.e.* TA sites) for each gene (30). Of note, 43 genes lacked unique TA insertion site
240 flanking sequences, consequently essentiality of these genes cannot be accurately
241 addressed; genes without unique TA flanking sequences are listed in Table S1.
242 Analysis revealed that 136 genes had no reads at all and 296 genes showed a log_2
243 fold-change (actual/expected sequence reads) cut-off below -6.86. Next, to reduce
244 the number of false-positive identified essential genes, we applied a 0.95
245 probability (calculated with a derivative of the Poisson law) cut-off that the gene, if
246 possible, was inactivated by a Tn insertion (based on 56,176 unique Tn mutants).
247 Application of this cut-off excluded 37 genes from the analysis, yielding a total of
248 395 genes that were found to be essential for *in vitro* growth and survival in TY
249 medium of *H. seropedicae* SmR1 (Table S2).

250 Essential genes were relatively uniformly distributed across the genome.
251 However, eight regions larger than 100,000 bp were found to be dispensable for
252 growth and survival. The two largest dispensable regions were located between
253 Hsero_2418 and Hsero-4580 (*trnL*) (202,525 bp), and Hsero_4426 (*glmS*) and
254 Hsero_4580 (194,479 bp).

255 **In-depth analysis of the genes required for *in vitro* growth and survival.**
256 Out of the 395 genes identified as being required for growth and survival in TY
257 medium, 22 corresponded to transfer RNA genes and one gene corresponded to a

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

258 23S ribosomal RNA gene (Hsero_4734, *rrlC*) (Table S2). The other two 23S
259 ribosomal RNA genes *rrlA* (Hsero_0480) and *rrlB* (Hsero_3882) could not be
260 evaluated for their essentiality as they had no unique TA insertion site flanking
261 sequence (*rrlA*), or the probability of inactivation was only 0.632 (*rrlB*). From the
262 remaining 372 protein coding genes required for *in vitro* growth, 346 were assigned
263 a Cluster of Orthologous Groups (COG) identifier. The COG categories
264 significantly enriched among the genes identified to be essential in *H. seropedicae*
265 are shown in Fig. 2B, and were: "cell cycle control, cell division, chromosome
266 partitioning" [D] "nucleotide transport and metabolism" [F], "coenzyme transport
267 and metabolism" [H], "translation, ribosomal structure and biogenesis" [J],
268 "replication, recombination and repair [L] and "cell wall/membrane/envelope
269 biogenesis" [M]. COG category "RNA processing and modification" has only one
270 representative, the product of the gene Hsero_1434, which is predicted to encode
271 for an oligoribonuclease, and is the only gene belonging to COG category A in the
272 *H. seropedicae* genome. A total of 1,624 protein coding genes containing
273 transmembrane domains or signal peptides are present in the genome and we
274 identified 72 of these as being essential, 64 are assigned to one or more COG
275 categories. As expected, the most represented COG category in this subset is "cell
276 wall/membrane/envelope biogenesis" [M].

277 **Essential metabolic pathways.** *In silico* analysis has revealed that *H.*
278 *seropedicae* cannot utilise L-histidine, L-arginine or L-lysine as carbon sources (8,
279 34). The L-histidine and L-lysine degradation pathways are incomplete and no
280 specific L-arginine transporter has been identified. In agreement with these
281 findings, our Tn-seq data indicates that the genes involved in the biosynthesis

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

282 pathways of these proteinogenic amino acids are essential. But also, and to our
283 knowledge not previously reported, both serine and glutamine synthesis seem to
284 be essential for *H. seropedicae* growth in TY medium. In the case of glutamine,
285 *glnA* (encoding glutamine synthetase) appears to be essential. Together with the
286 glutamine oxoglutarate aminotransferase (GOGAT enzyme), GlnA is the main
287 route of assimilation of NH_4^+ in bacteria (35, 36) and considering TY medium as a
288 nitrogen rich medium, we assume that GlnA activity should be low (36), and
289 therefore non-essential in this condition. GlnA activity and *glnA* expression has
290 previously been shown to be reduced, but not absent, when nitrogen is in excess
291 of 20 mM NH_4^+ (37). No reduction in expression of this gene or activity of the
292 enzyme was observed in the presence of glutamate (37). It is possible that nitrogen
293 may be more abundant, from amino acids and peptides, in TY medium, hence *glnA*
294 is probably expressed and GlnA active.

295 Another candidate essential gene related to nitrogen metabolism is *ntrX*
296 (Hsero_0069), which encodes a two-component response regulator protein.
297 Interestingly, a comparative genomics study reported that this gene is over-
298 represented in endophyte genomes compared to phytopathogens and rhizospheric
299 bacteria (4).

300 We determined as essential, several genes encoding proteins in the pentose
301 phosphate and glycolysis pathways. Three enzymes of the citric acid (TCA) cycle
302 were essential: aconitate hydratase (*acnA*, Hsero_2979), 2-oxoglutarate
303 dehydrogenase component E1 (*sucA*, Hsero_2969) and E3 (*lpdA*, Hsero_2967)
304 and two subunits of the succinate dehydrogenase (*sdhBC*, Hsero_2972 and
305 Hsero_2974). Hsero_2971 (annotated as hypothetical) was also found to be

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

306 essential, this gene has homology to *sdhE*, the product of which assists in the
307 covalent attachment of Flavin adenine dinucleotide (FAD) to SdhA (38), the
308 product of the gene Hsero_2973, which was not identified as essential.

309 Functional redundancy between genes precludes essentiality of central
310 metabolic pathways, however two homologous genes are not always redundant in
311 their functions. In the case of the already mentioned *acnA* gene (Hsero_2979)
312 which codes for the TCA cycle enzyme aconitate hydratase, *H. seropedicae*
313 contains in its genome another gene annotated as *acnA* (Hsero_2283) with a
314 41.89% identity. However, a mutant in this gene was identified in a Tn mutant
315 library previously described in the closely related strain *H. seropedicae* Z67 (39).
316 This suggests that *acnA* Hsero_2283 does not participate in the TCA cycle.
317 Homologs of *iscA* (Hsero_3845 and Hsero_3142), a gene involved in Fe-S cluster
318 biogenesis, were identified as essential genes. The two genes belong to the same
319 COG (00316), pFAM (01521) and TIGRFAM (00049) families. Their essentiality
320 indicates that they are not functionally redundant. This suggests the existence of
321 different Fe-S biogenesis machineries for different proteins.

322 The *hfq* gene (Hsero_2948), encoding an RNA chaperone, is also essential
323 for *H. seropedicae* SmR1 in the conditions studied. Several attempts to construct a
324 defined deletion mutant in this gene were unsuccessful (Emmanuel de Souza,
325 personal communication). The gene Hsero_4268 encodes a plasmid maintenance
326 system antidote protein which we identified as being essential in our analysis.
327 RNA-seq expression analysis showed this gene and its toxin counterpart gene,
328 Hsero_4269 were actively expressed in minimal media (13), and indicates that
329 there is an active toxin-antitoxin system in *H. seropedicae* SmR1.

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

330 **Critical reflection on identified candidate essential genes.** In this study,
331 the Tn mutants are grown in pools, consequently Tn mutants with a reduced fitness
332 (*i.e.* slowly growing/dividing bacteria) will be present at a lower abundance in the
333 pool (reflected by a lower read count of Tn flanking sequences), corresponding
334 genes can be tagged as essential in our analysis, *i.e.* the number of sequence
335 reads per gene falls below the essentiality cut-off (21).

336 As part of our preliminary studies into the Tn libraries, we performed Sanger
337 sequencing to identify the Tn insertion site in eight randomly selected mutants.
338 Through this, we identified a mutant in which the Tn was inserted in the *dadX* gene
339 (Hsero_2150). The enzyme encoded by this gene is predicted to catalyze the
340 conversion of L-alanine to D-alanine, which then is incorporated into the
341 peptidoglycan biosynthesis pathway by the D-alanine-D-alanine ligase protein,
342 encoded by the gene *ddlB* (Hsero_0338). Interestingly, according to our Tn-seq
343 data, *dadX* appears to be essential in *H. seropedicae* (Table S2). We hypothesised
344 that D-alanine may be synthesised *via* an alternative pathway at a lower rate,
345 allowing the recovery of the mutant as a single colony but not after growth of a
346 pooled Tn mutant, during which there is competition between Tn mutants. We
347 hypothesized that the alternative pathway could rely on Hsero_4778, which is
348 predicted to encode D-alanine transaminase [EC:2.6.1.21], which catalyzes the
349 interconversion of pyruvate and D-glutamate to D-alanine and 2-oxoglutarate.

350 **Comparative analysis of candidate essential genes with other bacteria.**
351 To identify orthologs of the 372 (including *dadX*) protein-encoding candidate
352 essential genes in *H. seropedicae*, a BlastP search was performed (*E*-value cut-off
353 of 1×10^{-5}, > 30% sequence identity over > 50% of the sequence length) against

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

354 essential genes in 39 bacterial strains from 28 different bacterial species present in
355 the Database of Essential Genes (DEG) (database accessed in July 2016) (40). A
356 total of 347 *H. seropedicae* SmR1 essential genes had at least one essential
357 ortholog in the bacterial species present in the DEG. The 347 *H. seropedicae*
358 SmR1 genes had 8,472 orthologs in the database (Table S3). The high percentage
359 of genes identified as essential in our study, that were also described as being
360 essential in other bacterial species, reinforces the quality of our candidate essential
361 gene set.

362 A total of 25 genes were uniquely essential in *H. seropedicae* SmR1, *i.e.* no
363 orthologs were found to be essential in the DEG (Table S4). Of the 20 essential
364 proteins annotated as hypotheticals, 14 are only essential in *H. seropedicae*. Three
365 are proteins related to secretion systems; Hsero_0751 and Hsero_0943 are related
366 to the type VI secretion system and Hsero_0804 is related to the Type III secretion
367 system of *H. seropedicae*. Type VI secretion systems are important for bacterial
368 competition through contact-dependent killing if competitors [41]. RNAseq analysis
369 of *H. seropedicae* grown in minimal media, or attached to maize roots, showed that
370 genes encoding the Type III secretion system were not expressed in either
371 condition (13). This might suggest that Hsero_0804 could be conditionally-
372 essential, *i.e.* when the bacteria are grown in nutrient-rich media.

373 Five of the genes uniquely essential in *H. seropedicae* encode for
374 transcriptional regulators, three of them belonging to the "transcription" COG
375 category. Hsero_1027 is homologous to the global regulator gene *pecS* from the
376 phytopathogen *Dickeya dadantii* 33937, which is reported to repress the premature
377 expression of virulence genes during the first stage of plant infection, when *D.*

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

378 *dadantii* has to colonise the plant apoplast without provoking symptoms (42). A *D.*
379 *dadantii pecS* mutant is hypervirulent (43). The expression of *pecS* is down-
380 regulated (fold change of -12.24, *P*-value 9.39 x 10^{-9}) in *H. seropedicae* attached to
381 maize roots, implying that the genes repressed by PecS are expressed and maybe
382 important in this condition (13). However, products of these genes may be toxic
383 whilst expressed under nutrient-rich conditions. The genes Hsero_1086,
384 Hsero_2104 and Hsero_2356 encode for transcriptional regulators with Lambda
385 repressor-like, DNA binding domains. Hsero_2356 is part of a locus (Hsero_2351
386 to Hsero_2371) that has a lower GC content (56% GC) compared to the rest of the
387 SmR1 genome (63% GC). Interestingly, RNAseq expression profiling of bacteria
388 grown in minimal media as well as bacteria that are attached to maize roots
389 showed that genes of this locus (Hsero_2351 to Hsero_2356) were highly
390 expressed while the downstream genes of this genomic locus were not (13). We
391 also hypothesise that essentiality of these three regulators could be due to
392 repression of genes that could be lethal in the growth conditions used in our study.
393 The gene Hsero_4425 is annotated as a member of the AsnC family transcription
394 regulator proteins. It is divergently transcribed from the essential gene *glmS*
395 (Hsero_4426). Homologos to *glmS* have been described as essential in 25 other
396 bacterial species, and the arrangement of these two genes is conserved in many
397 Proteobacteria (data not shown). It is possible that that essentiality of Hsero_4425
398 in *H. seropedicae* SmR1 is related to the expression of *glmS*. Finally, the essential
399 hypothetical genes Hsero_2418 and Hsero_3074 are both adjacent to genes
400 encoding for homologs of the RNA polymerase sigma-E factor protein RpoE
401 (Hsero_2419 and Hsero_3073). Both genes have predicted transmembrane

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

402 helices and in the case of Hsero_2418, it belongs to the family Pfam:PF13490, a
403 putative zinc-finger found in several anti-sigma factor proteins. Homologs of these
404 two genes are always linked to RNA polymerase sigma factors in other bacteria.
405 We hypothesise that essentiality of these genes in TY medium could be due to
406 regulation of genes activated by the cognate sigma factors.

407 *H. seropedicae* candidate essential genes with described essential orthologs
408 in only one or two of the strains preset in the DEG are indicated in Table S3.
409 Interestingly, the gene Hsero_4295, which encodes for an outer membrane porin,
410 has essential orthologs only in the two β-Proteobacteria *Burkholderia thailandensis*
411 E264 and *B. pseudomallei* K96243 (44, 45), for which their essential gene set has
412 been described. Further, gene Hsero_4295 was reported to be up-regulated when
413 *H. seropedicae* attaches to wheat roots, but down-regulated when *H. seropedicae*
414 attached to maize roots (12, 13), suggesting that this gene may be involved in
415 host-specificity. Six of the genes described in Table S3 were found to be essential
416 only in *H. seropedicae* and in the soil inhabitant *B. thailandensis.* This subset of
417 genes might indicate essential systems for *Burkholderiales*.

Accepted Manuscript Posted Online
Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology
Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

418 **CONCLUSION**

419 In this study we have developed functional genomic techniques and resources for
420 the model endophyte *H. seropedicae*, not previously used before in this species or
421 in other bacterial endophytes. We have generated large, comprehensive Tn
422 libraries and we have characterised the Tn insertion sites using next generation
423 sequencing (Tn-seq). These nearly saturated Tn libraries allowed us to perform
424 robust essentiality analysis, and the results obtained are consistent with those
425 reported for other bacteria. Our analysis of *H. seropedicae* Tn libraries grown in TY
426 media has enabled us to define the essential genes under these growth conditions.
427 The obtained results enabled us to describe, at a functional level, the mechanisms
428 for growth of *H. seropedicae*, including synthetic pathways, toxins and regulatory
429 mechanisms. Furthermore, these Tn libraries represent a valuable resource for the
430 endophyte research community and will facilitate studies into the comprehensive
431 assessment of the genetic mechanisms of the endophytic lifestyle of *H.*
432 *seropedicae, i.e*. attachment to the root surface, internal colonisation of the plant
433 and the survival of the bacteria inside plants.

Accepted Manuscript Posted Online

434 **ACKNOWLEDGEMENTS**

435 *H. seropedicae* SmR1 was kindly provided by Prof Ray Dixon. Plasmid pSAM_R1
436 was kindly provided by Prof Chris Yost. We thank Dr Roy Chaudhuri for sharing the
437 read count mapping Perl script. We thank Dr Aldert Zomer for providing R scripts
438 for the rarefaction analysis.

439

440 **FUNDING INFORMATION**

441 This work was funded by an EMBO ASTF 425-2014 short-term fellowship awarded
442 to FR and conducted in the laboratory of AJG. SPWdV and AB were funded by
443 BBSRC grant BB/K004514/1. The authors have no conflicting financial interests.
444 The funders had no role in the study design, data collection and interpretation, or
445 the decision to submit the work for publication.

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Accepted Manuscript Posted Online
Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology
Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

446 **References**

447 1. **Edwards, J. Johnson, C. Santos-Medellin, C. Lurie, E. Podishetty, N.K.**
448 **Bhatnagar, S. Eisen, J.A. Sundaresan, V.** 2015. Structure, variation, and
449 assembly of the root-associated microbiomes of rice. Proc. Natl. Acad. Sci.
450 U.S.A. **112**: E911-920.

451 2. **Coleman-Derr, D. Tringe, S.G.** 2014. Building the crops of tomorrow:
452 advantages of symbiont-based approaches to improving abiotic stress
453 tolerance. Front. Microbiol. **5**: 283.

454 3. **Mendes, R. Garbeva, P. Raaijmakers, J.M.** 2013. The rhizosphere
455 microbiome: significance of plant beneficial, plant pathogenic, and human
456 pathogenic microorganisms. FEMS. Microbiol. Rev. **37**: 634-663.

457 4. **Hardoim, P.R. van Overbeek, L.S. Berg, G. Pirttila, A.M. Compant, S.**
458 **Campisano, A. Doring, M. Sessitsch, A.** 2015. The Hidden World within
459 Plants: Ecological and Evolutionary Considerations for Defining Functioning of
460 Microbial Endophytes. Microbiol. Mol. Biol. Rev. **79**: 293-320.

461 5. **Baldani, J. Baldani, V.L.D. Seldin, L. Döbereiner, J.** 1986. Characterization
462 of *Herbaspirillum seropedicae* gen. nov., sp. Nov., a root-asociated nitrogen-
463 fixing bacterium. Int. J. Syst. Bacteriol. **36**: 86-93.

464 6. **Tripplett, E.W.** 2007. Prospects for Significant Nitrogen Fixation in Grasses
465 From Bacterial Endophytes, p 303-314. In Elmerich C, Newton WE (ed),
466 Associative and Endophytic Nitrogen-fixing Bacteria and Cyanobacterial
467 Associations doi:10.1007/1-4020-3546-2_13. Springer Netherlands, Dordrecht.

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

468 7. **Berrtani, I. Abbruscato, P. Piffanelli, P. Subramoni, S. Venturi, V.** 2016.
469 Rice bacterial endophytes: Isolation of a collection, identification of beneficial
470 strains and microbiome analysis. Environ. Microbiol. Rep. **8**: 388-398.
471 8. **Pedrosa, F.O. Monteiro, R.A. Wassem, R. Cruz, L.M. Ayub, R.A. Colauto,**
472 **N.B. Fernandez, M.A. Fungaro, M.H. Grisard, E.C. Hungria, M. Madeira,**
473 **H.M. Nodari, R.O. Osaku, C.A. Petzl-Erler, M.L. Terenzi, H. Vieira, L.G.**
474 **Steffens, M.B. Weiss, V.A. Pereira, L.F. Almeida, M.I. Alves, L.R. Marin, A.**
475 **Araujo, L.M. Balsanelli, E. Baura, V.A. Chubatsu, L.S. Faoro, H. Favetti, A.**
476 **Friedermann, G. Glienke, C. Karp, S. Kava-Cordeiro, V. Raittz, R.T.**
477 **Ramos, H.J. Ribeiro, E.M. Rigo, L.U. Rocha, S.N. Schwab, S. Silva, A.G.**
478 **Souza, E.M. Tadra-Sfeir, M.Z. Torres, R.A. Dabul, A.N. Soares, M.A.**
479 **Gasques, L.S. Gimenes, C.C. Valle, J.S. Ciferri, R.R. Correa, L.C. Murace,**
480 **N.K. Pampile, J.A. Patussi, E.V. Prioli, A.J. Prioli, S.M. Rocha, C.L.**
481 **Arantes, O.M. Furlaneto, M.C. Godoy, L.P. Oliveira, C.E. Satori, D. Vilas-**
482 **Boas, L.A. Watanabe, M.A. Dambros, B.P. Guerra, M.P. Mathioni, S.M.**
483 **Santos, K.L. Steindel, M. Vernal, J. Barcellos, F.G. Campo, R.J. Chueire,**
484 **L.M. Nicolas, M.F. Pereira-Ferrari, L. Silva, J.L. Gioppo, N.M. Margarido,**
485 **V.P. Menck-Soares, M.A. Pinto, F.G. Simao Rde, C. Takahashi, E.K. Yates,**
486 **M.G. Souza, E,M.** 2011. Genome of *Herbaspirillum seropedicae* strain SmR1,
487 a specialized diazotrophic endophyte of tropical grasses. PLoS. Genet. **7**:
488 e1002064.
489 9. **Suwantarat, N. Adams, L.T.L. Romagnoli, M. Carroll, K.C.** 2015. Fatal case
490 of *Herbaspirillum seropedicae* bacteremia secondary to pneumonia in an end-

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Accepted Manuscript Posted Online

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

491 stage renal disease patient with multiple myeloma. Diagn. Microbiol. Infect.
492 Dis. **82**: 331-333.

493 10. **Ziga, E.D. Druley, T. Burnham, C.A.** 2010. *Herbaspirillum* species bacteremia
494 in a pediatric oncology patient. J. Clin. Microbiol. **48**: 4320-4321.

495 11. **Monteiro, R.A. Balsanelli, E. Wassem, R. Marin, A.M. Brusamarello-**
496 **Santos, L.C.C. Schmidt, M.A. Tadra-Sfeir, M.Z. Pankievicz, V.C.S. Cruz,**
497 **L.M. Chubatsu, L.S.** 2012. *Herbaspirillum*-plant interactions: microscopical,
498 histological and molecular aspects. Plant. Soil. **356**: 175-196.

499 12. **Pankievicz, V.C. Camilios-Neto, D. Bonato, P. Balsanelli, E. Tadra-Sfeir,**
500 **M.Z. Faoro, H. Chubatsu, L.S. Donatti, L. Wajnberg, G. Passetti, F.**
501 **Monteiro, R.A. Pedrosa, F.O. Souza, E.M.** 2016. RNA-seq transcriptional
502 profiling of *Herbaspirillum seropedicae* colonizing wheat (*Triticum aestivum*)
503 roots. Plant. Mol. Biol. **90**: 589-603.

504 13. **Balsanelli, E, Tadra-Sfeir, M.Z. Faoro, H. Pankievicz, V.C. de Baura, V.A.**
505 **Pedrosa, F.O. de Souza, E.M. Dixon, R. Monteiro, R.A.** 2015. Molecular
506 adaptations of *Herbaspirillum seropedicae* during colonization of the maize
507 rhizosphere. Environ. Microbiol. doi:10.1111/1462-2920.12887.

508 14. **Alberton, D. Muller-Santos, M. Brusamarello-Santos, L.C. Valdameri, G.**
509 **Cordeiro, F.A. Yates, M.G. de Oliveira Pedrosa, F. de Souza, E.M.** 2013.
510 Comparative proteomics analysis of the rice roots colonized by *Herbaspirillum*
511 *seropedicae* strain SmR1 reveals induction of the methionine recycling in the
512 plant host. J. Proteome. Res. **12**: 4757-4768.

513 15. **Malfanova, N. Lugtenberg, B.J.J. Berg, G.** 2013. Bacterial Endophytes: Who
514 and Where, and What are they doing there?, p 391-403, Molecular Microbial

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Accepted Manuscript Posted Online

515 Ecology of the Rhizosphere doi:10.1002/9781118297674.ch36. John Wiley &
516 Sons, Inc.
517 16. **Mitter, B. Petric, A. Shin, M.W. Chain, P.S. Hauberg-Lotte, L. Reinhold-**
518 **Hurek, B. Nowak, J. Sessitsch, A.** 2013. Comparative genome analysis of
519 *Burkholderia phytofirmans* PsJN reveals a wide spectrum of endophytic
520 lifestyles based on interaction strategies with host plants. Front. Plant. Sci. **4**:
521 120.
522 17. **Sessitsch, A. Hardoim, P. Doring, J. Weilharter, A. Krause, A. Woyke, T.**
523 **Mitter, B. Hauberg-Lotte, L. Friedrich, F. Rahalkar, M. Hurek, T. Sarkar, A.**
524 **Bodrossy, L. van Overbeek, L. Brar, D,. van Elsas, J.D. Reinhold-Hurek,**
525 **B.** 2012. Functional Characteristics of an Endophyte Community Colonizing
526 Rice Roots as Revealed by Metagenomic Analysis. Mol. Plant. Microbe.
527 Interact. **25**: 28-36.
528 18. **Balsanelli, E. Serrato, R.V. de Baura, V.A. Sassaki, G. Yates, M.G. Rigo,**
529 **L.U. Pedrosa, F.O. de Souza, E.M. Monteiro, R.A.** 2010. *Herbaspirillum*
530 *seropedicae rfbB* and *rfbC* genes are required for maize colonization. Environ.
531 Microbiol. **12**: 2233-2244.
532 19. **Rosconi, F, Trovero, M.F. de Souza, E.M. Fabiano, E.** 2015. Serobactins
533 mediated iron acquisition systems optimize competitive fitness of
534 *Herbaspirillum seropedicae* inside rice plants. Environ. Microbiol.
535 doi:10.1111/1462-2920.13202.
536 20. **Kwon, Y.M. Ricke, S.C. Mandal, R.K.** 2015. Transposon sequencing:
537 methods and expanding applications. Appl. Microbiol. Biotechnol. **100**: 31-43.

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology

Accepted Manuscript Posted Online

538 21. **Barquist, L. Boinett, C.J. Cain, A.K.** 2013. Approaches to querying bacterial
539 genomes with transposon-insertion sequencing. RNA. Biol. **10**: 1161-1169.
540 22. **van Opijnen, T. Camilli, A.** 2013. Transposon insertion sequencing: a new
541 tool for systems-level analysis of microorganisms. Nat. Rev. Microbiol. **11**: 435-
542 442.
543 23. **Miller, V.L. Mekalanos, J.J.** 1988. A novel suicide vactor and its use in
544 construction of insertion mutations: osmoregulation of outer membrane
545 proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. J.
546 Bacteriol. **170**: 2575-2583.
547 24. **Bolivar, F. Rodriguez, R.L. Greene, P.J. Betlach, M.C. Heyneker, H.L.**
548 **Boyer, H.W. Crosa, J.H. Falkow, S.** 1977. Construction and characterization
549 of new cloning vehicle. II. A multipurpose cloning system. Gene. **2:** 95-113.
550 25. **Perry, B.J. Yost, C.K.** 2014. Construction of a mariner-based transposon
551 vector for use in insertion sequence mutagenesis in selected members of the
552 *Rhizobiaceae*. BMC. Microbiol. **14:** 298.
553 26. **Sambrook, J. Russell, D.W.** 2001. Molecular Cloning. 3rd ed. Cold Spring
554 Harbor, NY: Cold Spring Harbor Laboratory press.
555 27. **Karlyshev, A.V. Pallen, M.J. Wren, B.W.** 2000. Single-primer PCR procedure
556 for rapid identification of transposon insertion sites. Biotechniques. **28:** 1078,
557 1080, 1082.
558 28. **van Opijnen, T. Lazinski, D.W. Camilli, A.** 2014. Genome-Wide Fitness and
559 Genetic Interactions Determined by Tn-seq, a High-Throughput Massively
560 Parallel Sequencing Method for Microorganisms. Curr. Protoc. Mol. Biol. **106**:
561 7.16 1-24.

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Accepted Manuscript Posted Online

562 29. **Burghout, P. Zomer, A. van der Gaast-de Jongh, C.E. Janssen-Megens,**
563 **E.M. Francoijs, K.J. Stunnenberg, H.G. Hermans, P.W.** 2013. *Streptococcus*
564 *pneumoniae* folate biosynthesis responds to environmental CO2 levels. J,
565 Bacteriol **195**: 1573-1582.

566 30. **Zomer, A. Burghout, P. Bootsma, H.J. Hermans, P.W.M. van Hijum, S.A.**
567 2012. ESSENTIALS: software for rapid analysis of high throughput transposon
568 insertion sequencing data. PLoS. One. **7**: e43012.

569 31. **Storey, JD. Tibshirani, R. 2003.** Statistical significance for genomewide
570 studies. Proc. Natl. Acad. Sci. U.S.A. **100**: 9440-9445.

571 32. **Carver, T. Thomson, N. Bleasby, A. Berriman, M. Parkhill J.** 2009.
572 DNAPlotter: circular and linear interactive genome visualization. Bioinformatics.
573 **25**: 119-120.

574 33. **Markowitz, V.M. Chen, I.M. Palaniappan, K. Chu, K. Szeto, E. Grechkin, Y.**
575 **Ratner, A. Anderson, I. Lykidis, A. Mavromatis, K. Ivanova, N.N. Kyrpides,**
576 **N.C.** 2010. The integrated microbial genomes system: an expanding
577 comparative analysis resource. Nucleic. Acids. Res. **38**: D382-390.

578 34. **Klassen, G. Pedrosa, F.O. Souza, E.M. Funayama, S. Rigo, L.U.** 1997.
579 Effect of nitrogen compounds on nitrogenase activity in *Herbaspirillum*
580 *seropedicae* SMR1. Can. J. Microbiol. **43**: 887-891.

581 35. **Suzuki, A. Knaff, D.B.** 2005. Glutamate synthase: structural, mechanistic and
582 regulatory properties, and role in the amino acid metabolism. Photosynth. Res.
583 83: 191-217.

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Accepted Manuscript Posted Online

584 36. **Chobatsu, L. Monteiro, R. Souza, E. Oliveira, M. Yates, M. Wasser, R.**
585 **Bonatto, A. Huergo, L. Steffens, M. Rigo, L. Pedrosa, Fd.** 2012. Nitrogen
586 fixation control in *Herbaspirillum seropedicae*. Plant. Soil. **356**: 197-207.
587 37. **Persuhn, D.C. Souza, E.M. Steffens, M.B. Pedrosa, F.O. Yates, M.G. Rigo,**
588 **L.U.** 2000. The transcriptional activator NtrC controls the expression and
589 activity of glutamine synthetase in *Herbaspirillum seropedicae*. FEMS.
590 Microbiol. Lett. **192**, 217-221.
591 38. **McNeil, M.B. Hampton, H.G. Hards, K.J. Watson, B.N. Cook, G.M. Fineran,**
592 **P.C.** 2014. The succinate dehydrogenase assembly factor, SdhE, is required
593 for the flavinylation and activation of fumarate reductase in bacteria. FEBS.
594 Lett. **58**8: 414-421.
595 39. **Rosconi, F. Souza, E.M. Pedrosa, F.O. Platero, R.A. Gonzalez, C.**
596 **Gonzalez, M. Batista, S. Gill, P.R. Fabiano, E.R.** 2006. Iron depletion affects
597 nitrogenase activity and expression of *nifH* and *nifA* genes in *Herbaspirillum*
598 *seropedicae*. FEMS. Microbiol. Lett. **258**: 214-219.
599 40. **Luo, H. Lin, Y. Gao, F. Zhang, C.T. Zhang, R.** 2014. DEG 10, an update of
600 the database of essential genes that includes both protein-coding genes and
601 noncoding genomic elements. Nucleic. Acids. Res. **42**: D574-580.
602 41. **Dong, T.G. Ho, B.T. Yoder-Himes, D.R. Mekalanos, J.J.** 2013. Identification
603 of T6SS-dependent effector and immunity proteins by Tn-seq in *Vibrio cholera*.
604 Proc. Natl. Acad. Sci. U.S.A. **110**: 2623-2628.
605 42. **Mhedbi-Hajri, N. Malfatti, P. Pedron, J. Gaubert, S. Reverchon, S. Van**
606 **Gijsegem, F.** 2011. PecS is an important player in the regulatory network

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

607 governing the coordinated expression of virulence genes during the interaction
608 between *Dickeya dadantii* 3937 and plants. Environ. Microbiol. **13:** 2901-2914.
609 43. **Hommais, F. Oger-Desfeux, C. van Gijsegem, F. Castang, S. Ligori, S.**
610 **Expert, D. Nasser, W. Reverchon, S.** 2008. PecS is a global requlator of the
611 symptomatic phase in phytopathogenic bacterium *Erwinia chrysanthemi* 3937.
612 J. Bacteriol. **190**: 75802-7522.
613 44. **Baugh, L. Gallagher, L.A. Patrapuvich, R. Clifton, M.C. Gardberg, A.S.**
614 **Edwards, T.E. Armour, B. Begley, D.W. Dieterich, S.H. Dranow, D.M.**
615 **Abendroth, J. Fairman, J.W. Fox, D. 3rd, Staker, B.L. Phan, I. Gillespie, A.**
616 **Choi, R. Nakazawa-Hewitt, S. Nguyen, MT. Napuli, A. Barrett, L. Buchko,**
617 **GW. Stacy, R. Myler, P.J. Stewart, L.J. Manoil, C. Van Voorhis, WC.** 2013.
618 Combining functional and structural genomics to sample the essential
619 *Burkholderia* structome. PLoS. One. **8:** e53851.
620 45. **Moule, M.G. Hemsley, C.M. Seet, Q. Guerra-Assuncao, J.A. Lim, J. Sarkar-**
621 **Tyson, M. Clark, T.G. Tan, P.B. Titball, R.W. Cuccui, J. Wren, B.W.** 2014.
622 Genome-wide saturation mutagenesis of *Burkholderia pseudomallei* K96243
623 predicts essential genes and novel targets for antimicrobial development.
624 MBio. **5:** e00926-00913.

AEM

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Applied and Environmental Microbiology

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

625 **FIGURE AND TABLE LEGENDS**

626 **FIG 1** Characterization of *H. seropedicae* SmR1 Tn mutant libraries. A) Cumulative
627 number of Tn insertions when adding the different constructed mutant libraries and
628 the number of unique Tn insertion mutants yielded. B) Rarefaction analysis of
629 intragenic Tn insertion positions indicates near saturation of the number of genes
630 that can be inactivated with a Tn. C) Circular genome visualisation indicating genes
631 required for growth and survival in *H. seropedicae* SmR1.

632

633 **FIG 2** Identification and characterisation of *H. seropedicae* SmR1 essential genes.
634 A) Density plot log2 fold-change measured reads/expected reads per gene. Dots
635 indicate gene essentiality cut-off values. B) Functional class enrichment analysis of
636 essential genes based on COG categories. Bars represent the number of essential
637 genes assigned to each COG category, with the number of essential genes given
638 over the total number of genes in the COG category displayed right to each bar.
639 COG enrichment was analysed using a Fisher-exact test and corrected for multiple
640 testing using Q-value (31), * *Q*, 0.1, ** *Q*, 0.01, *** *Q*, 0.001.

641

642 **Table 1** Bacterial strains and plasmids used in this study. (Amp: ampicillin, Km:
643 kanamycin, Tc: tetracycline).

644

645 **Table 2** Primer sequences used in this study.

646

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

647 **Table 3** Tn mutant libraries constructed in *H. seropedicae* SmR1 (Km; kanamycin,

648 Tc; tetracycline). Criteria for identification of unique Tn insertion sites are listed in

649 Materials and Methods.

650

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Accepted Manuscript Posted Online

651

652

Strains and plasmids	Relevant genotype/description	Source/reference
Herbaspirillum seropedicae		
SmR1	Wild type, spontaneous streptomycin resistant mutant of strain Z78	(5)
Escherichia coli		
NEB 5-alpha	Subcloning Efficiency™ DH5α™ derived competent cells. *fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs
TransforMax™ EC100D™ *pir*$^{+}$	Electrocompetent cells constituvely expressing the *pir* gene. *F mcrA Δ(mrr-hsdRMS-mcrBC) φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ- rpsL (Str*^{R}*) nupG pir*$^{+}$*(DHFR)*	Epicentre
SM10-λpir	Donor strain carrying the transfer genes of the broad host range IncP type plasmid RP4. Km^{R}, *thi-1 thr leu tonA lacY supE recA::RP4-2-Tc::Mu pir*	(23)
Plasmids		
pBR322	Cloning vector. Amp^{R}, Tc^{R}.	(24)
pMiniT	Cloning vector. Amp^{R}.	New England Biolabs
pFRC002	*tetA* gene from pBR322 cloned in pMiniT. Amp^{R}, Tc^{R}.	This work
pSAM_R1	Suicide mobilizable vector, Amp^{R}, Km^{R}, *himar1-C9*	(25)
pSAM_R5	*tetA* gene from pFRC002 cloned in pSAM_R1. Amp^{R}, Tc^{R}.	This work

653

654 **Table 1**

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Applied and Environmental Microbiology AEM Applied and Environmental Microbiology

Accepted Manuscript Posted Online

Primer name	Characteristics	Sequence 5'-3'
Tet_FW1_*XhoI*	For amplification of *tetA* gene	CTCGAGTCTCATGTTTGACAGCTTATCATCG
Tet_FW1_*XbaI*	For amplification of *tetA* gene	TCTAGAGTTTGCGCATTCACAGTTCTCCG
Km_RV2	For identification of Tn insertion sites by single primer PCR	CGTGCAATCCATCTTGTTCAATC
Oligos for Tn-seq		
PBGSF29 ATCACG	adapter primer with ATCACG barcode	TTCCCTACACGACGCTCTTCCGATCTATCACGNN
PBGSF30 ATCACG	adapter primer with ATCACG barcode	P-CGTGATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P
PBGSF29 CGATGT	adapter primer with CGATGT barcode	TTCCCTACACGACGCTCTTCCGATCTCGATGTNN
PBGSF30 CGATGT	adapter primer with CGATGT barcode	P-ACATCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P
PBGSF29 TGACCA	adapter primer with TGACCA barcode	TTCCCTACACGACGCTCTTCCGATCTTGACCANN
PBGSF30 TGACCA	adapter primer with TGACCA barcode	P-TGGTCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P
PBGSF29 CTTGTA	adapter primer with CTTGTA barcode	TTCCCTACACGACGCTCTTCCGATCTCTTGTANN
PBGSF30 CTTGTA	adapter primer with CTTGTA barcode	P-TACAAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P
PBGSF29 CGTACG	adapter primer with CGTACG barcode	TTCCCTACACGACGCTCTTCCGATCTCGTACGNN
PBGSF30 CGTACG	adapter primer with CGTACG barcode	P-CGTACGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P
PBGSF23	GSF amplification primer 1	CAAGCAGAAGACGGCATACGAAGACCGGGGACTTATCATCCAACCTGT
PBGSF31	GSF amplification primer 2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

Table 2

Applied and Environmental Microbiology
AEM
Applied and Environmental Microbiology
Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

659

Library (antibiotic marker)	Estimated Tn library size	Sequence reads	Aligned reads	Aligned reads (%)	Insertion site flanking sequences hit in library	% Insertion sites hit in library	Average reads per flanking sequence
A (Km)	24,000	14,008,866	10,465,200	74.7	26,590	15.5	394
B (Km)	55,000	3,046,565	2,491,660	81.8	19,038	11.1	131
C (Km)	90,000	25,122,546	22,284,725	88.7	52,327	30.5	426
D (Km)	140,000	20,353,571	18,537,039	91.1	50,639	29.5	366
E (Tc)	70,000	12,165,012	10,998,288	90.4	30,984	18.0	355
F (Tc)	50,000	7,775,781	6,969,727	89.6	28,492	16.6	245

660

661 **Table 3**

662

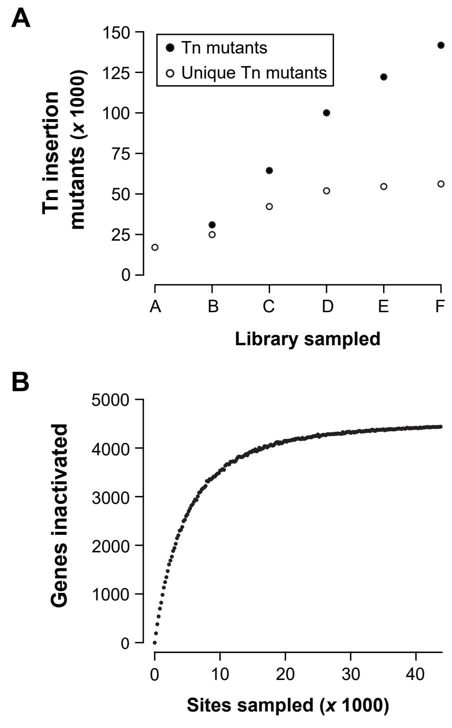
Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

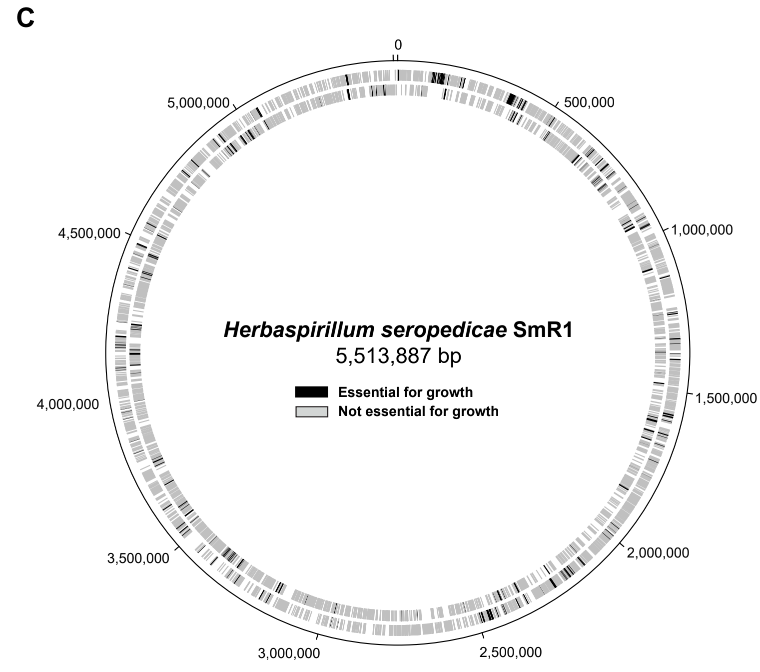
Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology





Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

Downloaded from http://aem.asm.org/ on October 26, 2016 by MRC LABORATORY OF MOLEC BIOLOG