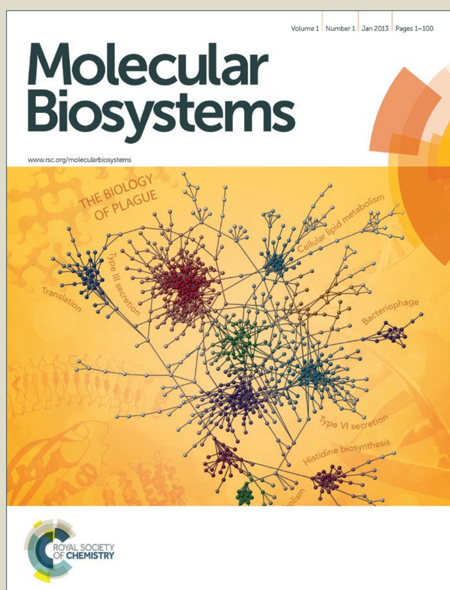
Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

www.rsc.org/molecularbiosystems

1 **Title:**

2 Transcriptional changes are involved in phenotype switching in *Streptococcus equi*

3 subspecies *equi*.

4

5 **Authors:**

6 Karen F. Steward[1*], Carl Robinson[1], and Andrew S. Waller[1].

7

8 **Affiliation:**

9 *[1]Animal Health Trust, Kentford, Newmarket, Suffolk, CB8 7UU, UK*

10

11 [*] Corresponding author:

12 E-mail: karen.steward@aht.org.uk, telephone: +44 (0)1638 750000, fax: +44 (0)1638

13 555643.

14

15 **Abstract:**

16 Phenotypic heterogeneity within a population of bacteria, through genetic or transcriptional

17 variation, enables survival and persistence in challenging and changing environments. We

18 report here that a recent clinical isolate of *S. equi*, strain 1691 (*Se*1691), yielded a mixture of

19 reduced capsule and mucoid colonies on primary isolation when grown on Colistin-oxolinic

20 acid blood agar (COBA) streptococcal selective plates. Passaging colonies of *Se*1691, with

21 a reduced capsule phenotype maintained this mixed phenotype. In contrast, passaging

22 mucoid colonies fixed the mucoid phenotype, suggesting adaptive genetic or transcriptional

23 changes in response to growth on artificial media. However, despite obvious phenotypic and

24 transcriptional differences, there were no apparent differences in the genome sequences of

25 *Se*1691 recovered from colonies with a mucoid or reduced capsule phenotype. We identified

26 105 differentially transcribed genes in the transcriptomes of reduced capsule and mucoid

27 colonies. The reduced capsule phenotype was associated with a significant reduction in

Molecular BioSystems Accepted Manuscript

28 transcription of the *has* locus (SEQ_0269 Q =0.0015, SEQ_0270 Q =0.0015, SEQ_0271 Q
29 =0.0285) and the amount of hyaluronic acid on the surface of *S. equi* recovered from non-
30 mucoid colonies ($P = 0.017$). Significant differences in the transcription of 21 surface and
31 secreted proteins were also observed. Our data show that changes in the bacterial
32 transcriptome are linked to the mixed colony phenotype of *Se*1691.
33

34 **Key words:**
35 *Streptococcus equi*, transcriptomics, phenotype, capsule.
36

Molecular BioSystems Accepted Manuscript

37 **Abbreviations:**

38	COBA	Colistin-oxolinic acid blood agar
39	*E. faecalis*	*Enterococcus faecalis*
40	gDNA	Genomic DNA
41	qPCR	Quantitative PCR
42	RT	Reverse transcription
43	*S. equi*	*Streptococcus equi*
44	*Se*1691	*S. equi* strain 1691
45	*Se*4047	*S. equi* strain 4047
46	TE	Tris-EDTA
47	THA	Todd Hewitt agar
48	UK	United Kingdom

49

Molecular BioSystems Accepted Manuscript

50 **Introduction:**

51 The presence of phenotypically variant, but genetically homogeneous individuals within a
52 bacterial population enable bacteria to swiftly take advantage of changing circumstances to
53 maximise infectivity, persistence and survival. Phenotypic changes are moderated by
54 alterations in gene transcription resulting in modified growth [1]. This bet hedging behaviour
55 has been observed in many bacterial species affecting a host of different gene families,
56 including pilus expression and opacity in *Streptococcus pneumoniae* (*S. pneumoniae*) [2, 3]
57 and daughter cells suited to long and short term starvation in *Sinorhizobium meliloti* [4], and
58 can be induced experimentally in *Pseudomonas fluorescens* [5].

59

60 Strangles in horses is caused by the Gram-positive Lancefield group C pathogen
61 *Streptococcus equi* subspecies *equi* (*S. equi*). The clinical signs of strangles, typified by
62 pyrexia, followed by abscessation of lymph nodes in the head and neck, were first reported
63 by Jordanus Ruffus in 1251 [6]. *S. equi* typically produces β-haemolytic mucoid colonies when
64 cultured on streptococcal selective colistin-oxolinic acid blood agar (COBA) plates.
65 Haemolysis is due to the production of the streptolysin S toxin [7, 8], whilst the mucoid
66 phenotype is dependent on the production of a hyaluronic acid capsule [8, 9]. Hyaluronic acid
67 is prevalent in host tissue, and its production by *S. equi* is associated with increased
68 resistance to phagocytosis by masking the bacterium from the host immune system [9, 10].
69 However, masking the cell surface may also obscure surface proteins which are important
70 for attachment and persistence. We noted that some isolates of *S. equi* recovered from
71 horses with strangles in the United Kingdom (UK) display a mixed phenotype of both mucoid
72 and reduced capsule colonies when grown on COBA. In this study we show that phenotypic
73 variation of a recent isolate of *S. equi* was not due to genomic change, but is instead
74 underpinned by significant differences in gene transcription.

75

Molecular BioSystems Accepted Manuscript

76 **Methods:**

77 **Bacterial strains and culture conditions**

78 A recent clinical isolate of *S. equi,* strain 1691 (*Se*1691), which was recovered from a nasal
79 swab from a pony with clinical signs of strangles in Nottinghamshire, UK on the 4th March
80 2013, was selected on the basis that it displayed a mixed colony phenotype. β-haemolytic
81 colonies of *Se*1691 were recovered from glycerol stocks following overnight growth on
82 COBA (bioMérieux), blood agar (bioMérieux) or Todd Hewitt agar (THA) (Oxoid) at 37 ^{o}C in
83 an atmosphere containing 5 % CO_2.

84

85 **Colony phenotypes on different media**

86 Mucoid or reduced capsule colonies were inoculated onto COBA, blood agar or THA and the
87 proportion of mucoid or reduced capsule colonies that were returned were recorded over the
88 course of six passages.

89

90 **Measurement of total hyaluronic acid**

91 *Se*1691 was grown overnight on COBA at 37 ^{o}C in an atmosphere containing 5 % CO_2.
92 Twenty colonies of each phenotype from three plates were re-suspended in triplicate in 0.2
93 mL of water. Hyaluronic acid was released by vigorous shaking with 0.4 mL of chloroform for
94 15 seconds. The mixtures were incubated at room temperature for 1 hour and then
95 centrifuged at 16000 x g. The aqueous phase from each preparation was removed and the
96 amount of capsule that was recovered was determined as described previously [11, 12]. DNA
97 was extracted from 1 µL of the original colony suspensions using 20 µL of Microlysis Plus
98 (Thistle Scientific) and the number of copies of *S. equi* DNA measured by quantitative PCR
99 (qPCR) [13]. The amount of hyaluronic acid was expressed as femtograms per DNA copy and
100 statistical significance determined by an unpaired student's T-test.

101

Molecular BioSystems Accepted Manuscript

102 **Calculation of average coccal chain length**

103 *Se*1691 was grown overnight on COBA at 37 ^{o}C in an atmosphere containing 5 % CO_2. A
104 single colony of each phenotype was gently resuspended in 10 µl water on a microscope
105 slide and Gram stained using a kit (Pro Lab) as per the manufacturer's instructions. Slides
106 were examined using oil immersion under 100 x magnification on a Zeiss Axioskop and eight
107 visual fields of each were captured using a Zeiss AxioCam ICc1 and AxioVision release
108 4.7.2 software. Coccal morphology was visually compared and the average chain length
109 calculated and tested for difference between the phenotypes using a two-sided Mann-
110 Whitney U test.

111

112 **Sequencing gDNA from mucoid and reduced capsule colonies**

113 *Se*1691 was inoculated onto COBA and grown overnight at 37 ^{o}C in an atmosphere
114 containing 5 % CO_2. Twenty mucoid or reduced capsule colonies were re-suspended in 200
115 µL Gram positive lysis solution, containing 250 units/mL mutanolysin, 2 x 10^6 units/mL
116 lysozyme and 30 µg/mL hyaluronidase and incubated for 1 hour at 37 °C to allow efficient
117 cell lysis. Genomic DNA (gDNA) was then purified using GenElute spin columns according
118 to the manufacturer's instructions (all Sigma). DNA purity and quantity were determined
119 using a Nanodrop ND1000 spectrophotometer (NanoDrop Technologies) and Qubit
120 fluorometer (Life Technologies). DNA libraries were prepared using the Nextera XT DNA
121 sample preparation kit (Illumina), dual labelled using compatible indexing primer pairs from
122 the Nextera XT index kit (Illumina), purified using Agencourt AMPure XP beads (Beckman
123 Coulter) and quantified using the KAPA library quantification kit for Illumina sequencing
124 platforms (KAPABiosystems) as per the relevant manufacturer's instructions. The libraries
125 were then sequenced on an Illumina MiSeq genome sequencer, which generated 250 bp
126 paired end reads. Following determination of raw sequencing reads, SNPs and indels were
127 identified compared to the reference genome *S. equi* strain 4047 (*Se*4047) (accession
128 number NC_012471) [7] on the MiSeq instrument using GATK unified genotyper. The gVCF

Molecular BioSystems Accepted Manuscript

129 files of the mucoid and reduced capsule gDNA were compared to identify detectable
130 differences in the SNPs and indels present in each strain compared to *Se*4047. The *has*
131 locus was studied in particular, as SNPs, indels and amplifications of the locus have been
132 linked to phenotypic differences when grown *in vitro*[14]. Unfiltered SNP data was also
133 examined to identify any minority variants which may be present. The depth of sequence
134 reads across each draft genome was visualised using the Integrated Genomics Viewer (IGV)
135 [15] which enabled graphical viewing of the results aligned to the *Se*4047 genome to identify
136 likely regions of amplification. These data have been deposited in the European Nucleotide
137 Archive under the accession numbers: ERS713829 (*Se*1691 mucoid colonies) and
138 ERS713830 (*Se*1691 reduced capsule colonies).
139
140 **Extraction of RNA**
141 *Se*1691 was inoculated onto COBA and grown overnight at 37 ^{o}C in an atmosphere
142 containing 5 % CO_2. Three biological replicates of a 1 mL suspension of 10 reduced capsule
143 colonies or 10 mucoid colonies of *Se*1691 were mixed with 2 mL per suspension of RNA
144 protect (Qiagen) and pelleted by centrifugation for 10 minutes at 5000 x g. Each pellet was
145 re-suspended in 200 µL of tris-EDTA buffer (TE) (Fluka) containing 3 mg of lysozyme
146 (Sigma) and 500 U of mutanolysin (Sigma). Total RNA was then extracted using RNeasy
147 mini, and DNase kits (all Qiagen) as per the manufacturer's instructions with the following
148 alterations. After incubation at room temperature for 45 minutes with repeated vortexing, 700
149 µL of RLT buffer from the RNeasy mini kit was added and the tube vortexed briefly. The lysis
150 mixture was transferred to a tube containing 0.05 g of acid washed glass beads (Sigma) and
151 vortexed constantly for 5 minutes. The lysis mixture was centrifuged for 10 seconds at 16100
152 x g and the supernatant removed for RNA extraction with an RNeasy mini kit. RNA purity
153 and quantity were determined using a Nanodrop ND1000 spectrophotometer (NanoDrop
154 Technologies) and Qubit fluorometer (Life Technologies). rRNA was depleted from the total
155 RNA preparation, using a Ribozero magnetic kit for Gram-positive bacteria (Epicentre) as

Molecular BioSystems Accepted Manuscript

156 per the manufacturer's instructions and the amount of RNA re-quantified using a Qubit
157 fluorometer.
158
159 **Library preparation, quantification and sequencing**
160 Libraries from reduced capsule or mucoid colonies were prepared from the rRNA depleted
161 RNA samples using the NEBNext Ultra directional RNA library preparation kit for Illumina
162 (NEB), purified using Agencourt AMPure XP beads (Beckman Coulter) and quantified using
163 the KAPA library quantification kit for Illumina sequencing platforms (KAPABiosystems) as
164 per the relevant manufacturer's instructions. The libraries were then sequenced on an
165 Illumina MiSeq genome sequencer, which generated 75 bp paired end reads. These data
166 have been deposited in the European Nucleotide Archive under the accession numbers:
167 (ERS657961, ERS657962, ERS657963, ERS657964, ERS657965 and ERS657966) and in
168 the ArrayExpress database under the accession number E-MTAB-4147.
169
170 **Transcriptome analysis**
171 Raw sequencing reads were aligned to the completed reference genome of *Se*4047, using
172 Bowtie 2 [16] and transcriptomes were reconstructed using Cufflinks [17]. The transcriptomes of
173 the three suspensions of reduced capsule colonies were then compared to the
174 transcriptomes of the three suspensions of mucoid colonies using Cuffdiff, which identifies
175 differentially transcribed transcripts and genes [18]. Alignments and comparisons were also
176 made using the unfinished *Se*1691 genome as a reference sequence (accession number:
177 ERS657960) to identify any gross changes relating to regions or genes that may be unique
178 to *Se*1691. The corrected average transcript values generated by Cuffdiff for the two
179 conditions (reduced capsule or mucoid) for each gene were then converted into fold
180 difference values in reduced capsule compared to mucoid colonies, with a positive value
181 indicating an increase in the transcription of a coding sequence in reduced capsule
182 compared to mucoid colonies. Cuffdiff generated Q values (corrected P values) and genes
183 that were differentially transcribed significantly in reduced capsule compared to mucoid

Molecular BioSystems Accepted Manuscript

184 colonies were identified using the cut-off of Q ≤ 0.05. The enrichment of functional classes of
185 genes represented in the differentially transcribed gene set was tested for using the Fisher's
186 exact or chi-square test as appropriate.
187
188 **Verification of transcriptional differences identified from transcriptome analysis**
189 100 ng of RNA per sample were used to synthesise cDNA by reverse transcription (RT) in
190 20 µL reactions with random hexamers using a Verso cDNA kit (Thermo Scientific) as per
191 the manufacturer's instructions. Transcription of SEQ_0269, SEQ_0270, SEQ_0271,
192 SEQ_0402, SEQ_0546, SEQ_0938, SEQ_0999, SEQ_1728, SEQ_1817, SEQ_2190 and
193 the housekeeping gene *gyrA* (SEQ_1170) was then quantified by qPCR. Reactions
194 contained 1 x Kapa SYBR fast qPCR mix (Anachem), 0.3 µM forward and reverse primers
195 (Table S1) and 6 µL of a 1/10 dilution of cDNA made up to 20 µL with ddH_2O and
196 thermocycled at 95 °C for 3 minutes followed by 40 cycles of 95 °C for 30 seconds, 60 °C 10
197 seconds then 95 °C for 15 seconds. A ramp step from 60 °C to 95 °C with SYBR reads
198 every 0.3 °C was performed to calculate the dissociation curves of products. No template
199 and no RT controls were included to confirm the absence of contaminating DNA and RNA in
200 samples. Copy numbers were calculated from standard curves and normalised to *gyrA,*
201 which showed no significant difference in transcription between the phenotypes. The mean
202 values for the three biological replicates and the fold changes in gene transcription of
203 reduced capsule compared to mucoid colonies were calculated. Data were analysed by
204 unpaired student's T-test using independent samples and compared to the Q values
205 generated from analysis of the transcriptomics data.
206
207

Molecular BioSystems Accepted Manuscript

208 **Results and Discussion:**

209 **Phenotypic variation of *S. equi* isolates**

210 Colonies of *S. equi* following overnight growth on COBA are typically mucoid in appearance
211 through the production of an anti-phagocytic hyaluronic acid capsule [9, 10]. We noted that an
212 isolate of *S. equi*, *Se*1691, displayed a mixed phenotype of both mucoid and reduced
213 capsule colonies when grown on COBA (Figure 1A). The amount of hyaluronic acid
214 extracted from mucoid colonies was significantly greater than the amount recovered from
215 reduced capsule colony variants ($P = 0.017$) (Figure 1B), suggesting that this phenotype
216 may be due to reduced hyaluronic acid production and/or increased degradation through the
217 activity of hyaluronate lyase.
218
219 Inoculation of mucoid colonies of *Se*1691, on COBA yielded 100 % mucoid colonies
220 following overnight growth. In contrast, the inoculation of COBA with reduced capsule
221 colonies of *Se*1691 yielded a proportion of reduced capsule colonies (12 %, n = 6 passages
222 (95 % CI +/- 5 %)) and a majority of mucoid colonies following overnight growth. However,
223 inoculation of THA or blood agar with mucoid or reduced capsule colonies only yielded
224 mucoid colonies. Therefore, the reduced capsule phenotype of *Se*1691 was only apparent in
225 a minority of colonies following growth on COBA and our data suggest that the mucoid
226 colony phenotype was selected in vitro.
227

228 **Phenotypic variation of colonies was not due to mutation of the *S. equi* genome**

229 SNPs, indels and amplifications in the *has* locus of some isolates of *S. equi* recovered from
230 persistently infected horses have previously been identified as a source of phenotypic
231 variation, with multiple genetic variants exhibiting differing phenotypes being recovered from
232 the same horse[14]. The amplification of the *has* locus leads to the increased production of
233 hyaluronic acid and deletions within the locus to reduced hyaluronic acid production[14].
234 Glycerol stocks of *Se*1691 were created from a single pure colony from which mucoid and

Molecular BioSystems Accepted Manuscript

235 reduced capsule colonies were recovered. No SNPs, indels or amplifications of loci,
236 including the *has* locus, were identified that differentiated DNA from mucoid or reduced
237 capsule colonies. From analysing the unfiltered SNP data, even where a low quality SNP or
238 indel was present in a minority of reads, this was the case in the read data from both
239 phenotypes and most likely results from regions where sequence assembly was confounded
240 due to similarity with other regions of the *S. equi* genome. Our data demonstrate that the
241 phenotypic differences of *Se*1691 colonies were not due to mutation of the genome.
242
243 **Transcriptional differences exist between colony phenotypes of *S. equi***
244 To investigate if the variation in colony morphology seen on COBA was linked to the
245 regulation of gene transcription, we determined the transcriptomes of three reduced capsule
246 colony samples and three mucoid colony samples of *Se*1691. The mean number of reads
247 per sample was 11,299,173 for the reduced capsule phenotype and 6,526,730 for the
248 mucoid phenotype. Sequencing data with a depth of at least 10 - fold covered 82.9 % and
249 79.4 % of the genome respectively, 88.2 % of which was coding, with coverage of 2,059 and
250 2,049 of the 2,137 predicted coding sequences, respectively. One hundred and five genes
251 were differentially transcribed between the two colony phenotypes with Q values of ≤ 0.05
252 (Figure 2 and Table S2). In agreement with the reduced quantity of hyaluronic acid
253 recovered from colonies with a reduced capsule phenotype, the genes SEQ_0269 (*hasA*),
254 SEQ_0270 (*hasB*) and SEQ_0271 (*hasC*), which make up the *has* locus, responsible for the
255 production of the hyaluronic acid capsule, had significantly reduced transcription in the
256 reduced capsule compared to mucoid colonies (Figure 2). Genes for capsule synthesis were
257 enriched significantly in the differentially transcribed gene set ($P = 0.002$). In contrast, no
258 significant difference in the transcription of the prophage-encoded hyaluronate lyase,
259 SEQ_2045, which breaks down hyaluronic acid[19], was observed. Our data indicate that the
260 lack of the mucoid phenotype seen in the reduced capsule colonies of *Se*1691 is a result of
261 reduced hyaluronic acid capsule synthesis rather than increased breakdown.
262

Molecular BioSystems Accepted Manuscript

263 The greatest fold increase in transcription in reduced capsule compared to mucoid colonies
264 was seen in SEQ_0947 (16.56 - fold). SEQ_0947 is a conserved hypothetical protein
265 contained within an operon of unknown function, so the importance of this is unclear. The
266 greatest fold decrease was seen in SEQ_1291 (-23.48 - fold), which encodes a putative N-
267 acetylglucosaminidase family protein. In *Enterococcus faecalis* (*E. faecalis*), a protein of this
268 family is known to be involved with digestion of the septum and cell separation during
269 bacterial replication. Deletion of *atlA*, the gene that encodes the N-acetylglucosaminidase
270 family protein in *E. faecalis*, was shown to produce longer chains of cocci and alter the cell
271 shape [20]. However, in *S. equi* there was no change in the appearance of bacterial cells
272 recovered from mucoid or reduced capsule colonies or significant change in bacterial chain
273 length (mucoid mean chain length = 1.37, reduced capsule mean chain length = 1.46, P =
274 0.227, when the number of cells per chain across eight fields of view were measured which
275 contained an average of 124 chains per field) associated with the decrease in SEQ_1291
276 transcription. This suggests that SEQ_1291 may have a different role in *S. equi* or that the
277 residual transcription of SEQ_1291 was sufficient for normal cell separation.
278
279 The transcription of twenty one surface and secreted proteins in the two colony phenotypes
280 was significantly different between the mucoid and the reduced capsule variant with both
281 classes over-represented significantly in the differentially transcribed gene set ($P = 3.65 \times$
282 10^{-8} and $P = 8.24 \times 10^{-5}$ respectively) (Figure 2). Surface and secreted proteins are known to
283 be important to bacterial invasion and colonisation through their effects on adhesion, tissue
284 damage and immune evasion[21]. Four of the eight surface protein-encoding genes that
285 showed increased transcription and none of the three that showed reduced transcription
286 encoded surface anchored collagen-like proteins (SEQ_0090, SEQ_0280, SEQ_0855 and
287 SEQ_1817). The precise role of the collagen-like proteins produced by *S. equi* is not well
288 understood, but they have been implicated in invasion and evasion of the host immune
289 system in *S. pyogenes* [14, 22-26] and likely influence the ability of *S. equi* to interact with its
290 host. Five (SEQ_0256, SEQ_0402, SEQ_0721 (*eag*), SEQ_0999 (*ideE*), SEQ_0938 (*ideE2*)

Molecular BioSystems Accepted Manuscript

291 of the seven genes encoding the components of a recombinant multicomponent subunit
292 vaccine that protected six of seven vaccinated ponies from challenge with a virulent strain of
293 *S. equi* [27], showed significantly altered transcription. The immunogenic surface protein
294 SEQ_2190, which is a diagnostic target used in an indirect enzyme-linked immunosorbent
295 assay to identify horses that have been exposed to *S. equi* [28], showed increased
296 transcription in reduced capsule colonies. A significant reduction in the transcription of
297 SEQ_0546 (*sagA*), which encodes the streptolysin S precursor of *S. equi* [8], and SEQ_1728
298 (*seeL*), encoding the superantigen SeeL [29], was also identified in reduced capsule colonies.
299
300 The transcription of eleven putative transcriptional regulators, a significant proportion of the
301 regulators in the *S. equi* genome ($P = 0.0007$), was significantly lower in the reduced
302 capsule colonies (Table S2). These regulators include Mga-like, RofA-like, GntR-like, LysR-
303 like and MntR-like regulators. However, little is known about their activity in *S. equi*. The
304 transcription of SEQ_1661, which encodes the putative MntR metal dependent repressor of
305 the *mtsABC* locus, was 2-fold lower in the reduced capsule colony variants. The production
306 of the MtsABC metal transporter is important for the virulence of *Streptococcus uberis* [30],
307 which is closely related to *S. equi* [31]. However, no corresponding effect on the transcription
308 of the *mtsABC* locus was identified. No genes that encode putative or known transcriptional
309 regulators had higher levels of transcription in the reduced capsule variant colonies.
310
311 **Validation of transcriptomics data from *S. equi***
312 To validate the differential transcription of genes identified using the transcriptomics method,
313 qPCR was carried out on cDNA from reduced capsule and mucoid colonies in triplicate to
314 quantify the transcription of a sample of ten genes (Table 1). The qPCR assays confirmed
315 that the direction of the difference in gene transcription between reduced capsule and
316 mucoid colonies was the same by qPCR as it was by using transcriptomics for all of the ten
317 genes selected. Eight of the ten genes were confirmed to be significantly differentially
318 transcribed, while the transcription of the remaining two genes approached statistical

Molecular BioSystems Accepted Manuscript

319 significance (Figure 3). These data confirm that the use of transcriptomics for *S. equi* is a
320 valid method to identify differential gene transcription and suggests that the *ex vivo*
321 transcription profile of *S. equi* could be used to identify genes that may be important for this
322 organism to cause disease relative to the array of genes that are transcribed during *in vitro*
323 growth.
324
325

326 **Conclusions:**

327 The selective pressures on recent clinical isolates of *S. equi* when cultured on COBA rather
328 than blood agar alone, promotes a mixed phenotype of classical mucoid and reduced
329 capsule colonies. The mucoid but not reduced capsule phenotype becomes fixed following
330 passaging of colonies on COBA, which suggested the acquisition of a compensatory SNP,
331 amplification or indel. However, sequencing of colony material of both phenotypes
332 demonstrated that the phenotypes were linked to transcriptional, but not genetic variation.
333

334 The profile of genes that were differentially transcribed in the mucoid and reduced capsule
335 colony variants, suggest the involvement of a broad feedback effect that could play a role in
336 the adaptation of *S. equi* to the host environment or growth on artificial media. Many of the
337 genes identified here that were differentially transcribed have been linked to the virulence of
338 *S. equi* or the development of protective immunity. Our data shed new light on the regulation
339 of gene transcription in *S. equi* and reveal new genes the products of which, may contribute
340 to the virulence of *S. equi* and the effectiveness of protective vaccines against this endemic
341 disease.
342

343 **Acknowledgements:**

344 This work was funded by a grant from Petplan Charitable Trust (grant number S13-55).
345

Molecular BioSystems Accepted Manuscript

346 **Accession numbers for sequence data:**

347 Short reads for the *Se*1691 genome sequence (ERS657960), gDNA from mucoid
348 (ERS713829) and reduced capsule (ERS713830) colonies and RNA sequence from reduced
349 capsule (ERS657961, ERS657962 and ERS657963) or mucoid (ERS657964, ERS657965
350 and ERS657966) colonies have been deposited in the European Nucleotide Archive within
351 sequencing projects PRJEB9200 and PRJEB8454. RNA sequences have also been
352 deposited in the ArrayExpress database under the accession number E-MTAB-4147.
353

Molecular BioSystems Accepted Manuscript

354 **References:**

355 1. M. W. van der Woude, *FEMS Microbiol Lett*, 2006, **254**, 190-197.
356 2. A. Basset, K. H. Turner, E. Boush, S. Sayeed, S. L. Dove and R. Malley,
357 *Infect Immun*, 2011, **79**, 2974-2983.
358 3. J. O. Kim and J. N. Weiser, *J Infect Dis*, 1998, **177**, 368-377.
359 4. W. C. Ratcliff and R. F. Denison, *Curr Biol*, 2010, **20**, 1740-1744.
360 5. H. J. Beaumont, J. Gallie, C. Kost, G. C. Ferguson and P. B. Rainey, *Nature*,
361 2009, **462**, 90-93.
362 6. J. Rufus, ed., *De Medicina Equorum*, 1251.
363 7. J. Flanagan, N. Collin, J. Timoney, T. Mitchell, J. A. Mumford and N. Chanter,
364 *Microb Pathog*, 1998, **24**, 211-221.
365 8. C. Robinson, Z. Heather, J. Slater, N. Potts, K. F. Steward, D. J. Maskell, M.
366 C. Fontaine, J. J. Lee, K. Smith and A. S. Waller, *Vaccine*, 2015, **33**, 1160-
367 1167.
368 9. T. Anzai, J. F. Timoney, Y. Kuwamoto, Y. Fujita, R. Wada and T. Inoue, *Vet*
369 *Microbiol*, 1999, **67**, 277-286.
370 10. J. F. Timoney and P. Kumar, *Equine Vet J*, 2008, **40**, 637-642.
371 11. G. L. Darmstadt, L. Mentele, A. Podbielski and C. E. Rubens, *Infect Immun*,
372 2000, **68**, 1215-1221.
373 12. H. M. Schrager, J. G. Rheinwald and M. R. Wessels, *J Clin Invest*, 1996, **98**,
374 1954-1958.
375 13. K. Webb, C. Barker, T. Harrison, Z. Heather, K. F. Steward, C. Robinson, J.
376 R. Newton and A. S. Waller, *Vet J*, 2013, **195**, 300-304.
377 14. S. R. Harris, C. Robinson, K. F. Steward, K. S. Webb, R. Paillot, J. Parkhill, M.
378 T. Holden and A. S. Waller, *Genome Res*, 2015, **25**, 1360-1371.

Molecular BioSystems Accepted Manuscript

379 15. J. T. Robinson, H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander, G.
380 Getz and J. P. Mesirov, *Nat Biotechnol*, 2011, **29**, 24-26.
381 16. B. Langmead and S. L. Salzberg, *Nat Methods*, 2012, **9**, 357-359.
382 17. C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van
383 Baren, S. L. Salzberg, B. J. Wold and L. Pachter, *Nat Biotechnol*, 2010, **28**,
384 511-515.
385 18. C. Trapnell, D. G. Hendrickson, M. Sauvageau, L. Goff, J. L. Rinn and L.
386 Pachter, *Nat Biotechnol*, 2013, **31**, 46-53.
387 19. A. M. Lindsay, M. Zhang, Z. Mitchell, M. T. Holden, A. S. Waller, I. C. Sutcliffe
388 and G. W. Black, *Microbiology*, 2009, **155**, 443-449.
389 20. S. Mesnage, F. Chau, L. Dubost and M. Arthur, *J Biol Chem*, 2008, **283**,
390 19845-19853.
391 21. W. W. Navarre and O. Schneewind, *Microbiol Mol Biol Rev*, 1999, **63**, 174-
392 229.
393 22. C. C. Caswell, M. Barczyk, D. R. Keene, E. Lukomska, D. E. Gullberg and S.
394 Lukomski, *J Biol Chem*, 2008, **283**, 36168-36175.
395 23. C. C. Caswell, H. Oliver-Kozup, R. Han, E. Lukomska and S. Lukomski,
396 *FEMS Microbiol Lett*, 2009, **303**, 61-68.
397 24. H. A. Oliver-Kozup, M. Elliott, B. A. Bachert, K. H. Martin, S. D. Reid, D. E.
398 Schwegler-Berry, B. J. Green and S. Lukomski, *BMC Microbiol*, 2011, **11**,
399 262.
400 25. L. I. Pahlman, P. F. Marx, M. Morgelin, S. Lukomski, J. C. Meijers and H.
401 Herwald, *J Biol Chem*, 2007, **282**, 24873-24881.
402 26. Z. Yu, B. An, J. A. Ramshaw and B. Brodsky, *J Struct Biol*, 2014, **186**, 451-
403 461.

Molecular BioSystems Accepted Manuscript

404 27. B. Guss, M. Flock, L. Frykberg, A. S. Waller, C. Robinson, K. C. Smith and J.
405 I. Flock, *PLoS Pathog*, 2009, **5**, e1000584.
406 28. C. Robinson, K. F. Steward, N. Potts, C. Barker, T. A. Hammond, K. Pierce,
407 E. Gunnarsson, V. Svansson, J. Slater, J. R. Newton and A. S. Waller, *Vet J*,
408 2013, **197**, 188-191.
409 29. R. Paillot, C. Robinson, K. Steward, N. Wright, T. Jourdan, N. Butcher, Z.
410 Heather and A. S. Waller, *Infect Immun*, 2010, **78**, 1728-1739.
411 30. A. J. Smith, P. N. Ward, T. R. Field, C. L. Jones, R. A. Lincoln and J. A. Leigh,
412 *Infect Immun*, 2003, **71**, 4842-4849.
413 31. P. N. Ward, M. T. Holden, J. A. Leigh, N. Lennard, A. Bignell, A. Barron, L.
414 Clark, M. A. Quail, J. Woodward, B. G. Barrell, S. A. Egan, T. R. Field, D.
415 Maskell, M. Kehoe, C. G. Dowson, N. Chanter, A. M. Whatmore, S. D. Bentley
416 and J. Parkhill, *BMC Genomics*, 2009, **10**, 54.
417
418

Molecular BioSystems Accepted Manuscript

419 **Tables:**

420 **Table 1.** Genes selected from the transcriptomics analysis for analysis by qPCR, the fold
421 transcriptional change in reduced capsule relative to mucoid colonies and Q-value as
422 determined by transcriptome analysis are indicated. Full results are provided in Table S2.

423

Gene ID (name)	Role	Fold change	Q value
SEQ_0269 (*hasA*)	Hyaluronan synthase	-4.68	0.0015
SEQ_0270 (*hasB*)	UDP-glucose 6-dehydrogenase	-3.95	0.0015
SEQ_0271 (*hasC*)	UTP-glucose-1-phosphate uridylyltransferase 1	-2.21	0.0285
SEQ_0402	Sortase-processed surface protein	3.94	0.0056
SEQ_0546 (*sagA*)	Streptolysin S precursor	-4.60	0.0015
SEQ_0938 (*ideS*)	IgG endopeptidase	4.62	0.0015
SEQ_0999 (*ideE*)	IgG endopeptidase	2.88	0.0015
SEQ_1728 (*seeL*)	Superantigen	-4.71	0.0015
SEQ_1817 (*sclB*)	Collagen-like surface-anchored protein	4.15	0.0015
SEQ_2190	Sortase-processed surface protein	4.75	0.0015

424

425

Molecular BioSystems Accepted Manuscript

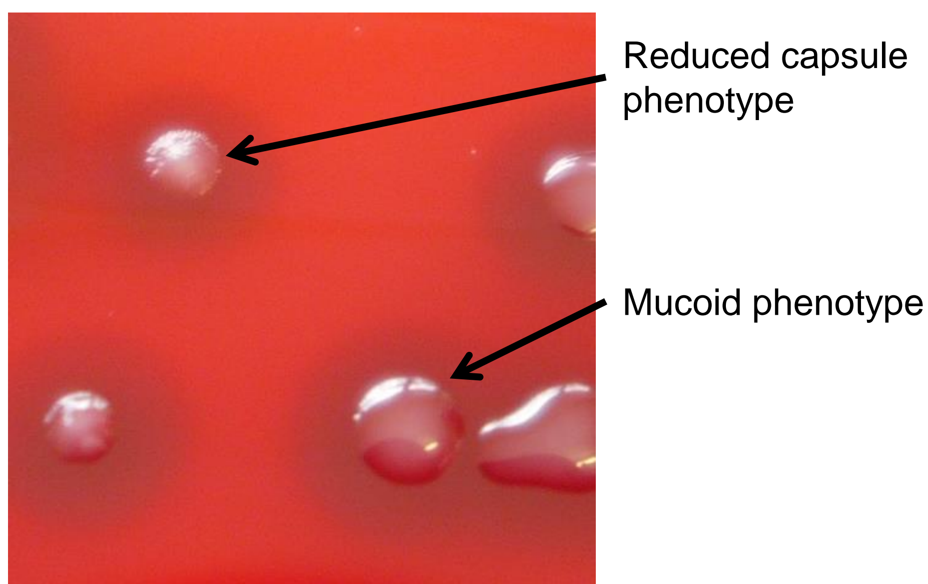
426 **Figures:**

427 **Figure 1.** A) Images of *S. equi* mucoid and reduced capsule colony phenotypes when grown

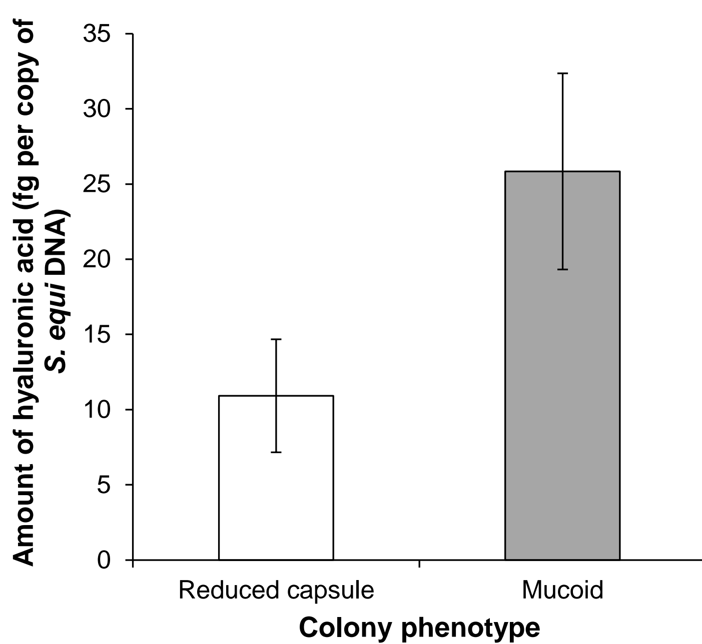
428 on COBA. B) Amount of hyaluronic acid recovered from mucoid and reduced capsule

429 colonies. Error bars indicate the 95 % confidence intervals.

A



B

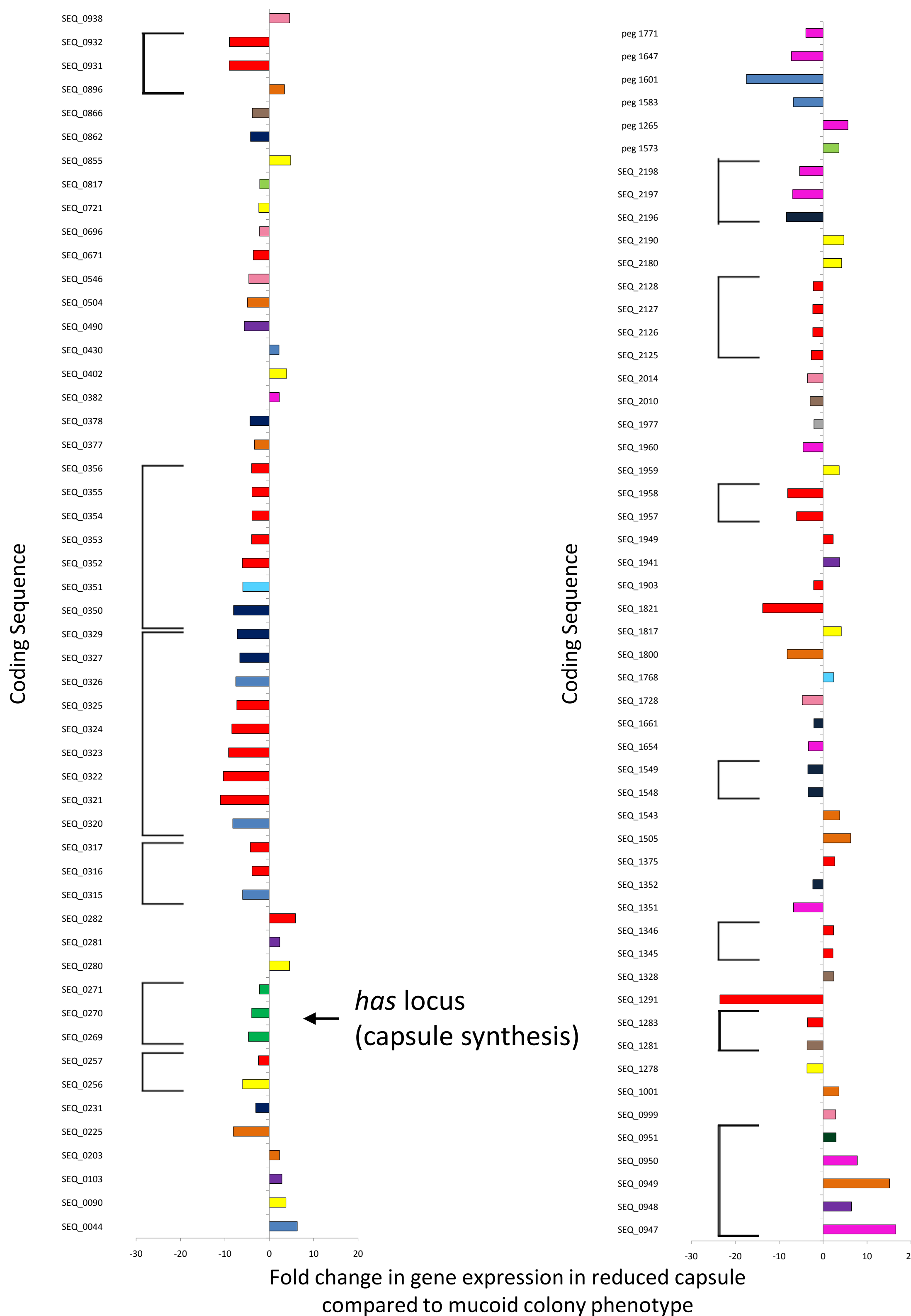


430

Molecular BioSystems Accepted Manuscript

431 **Figure 2.** Graph showing all the genes that were significantly differentially transcribed from
432 transcriptome sequencing ($Q \leq 0.05$) in reduced capsule compared to mucoid colonies of
433 *Se*1691. Genes that are in the same operon are indicated with a black bracket, functional
434 families are colour coded as indicated in the legend.
435

Molecular BioSystems Accepted Manuscript

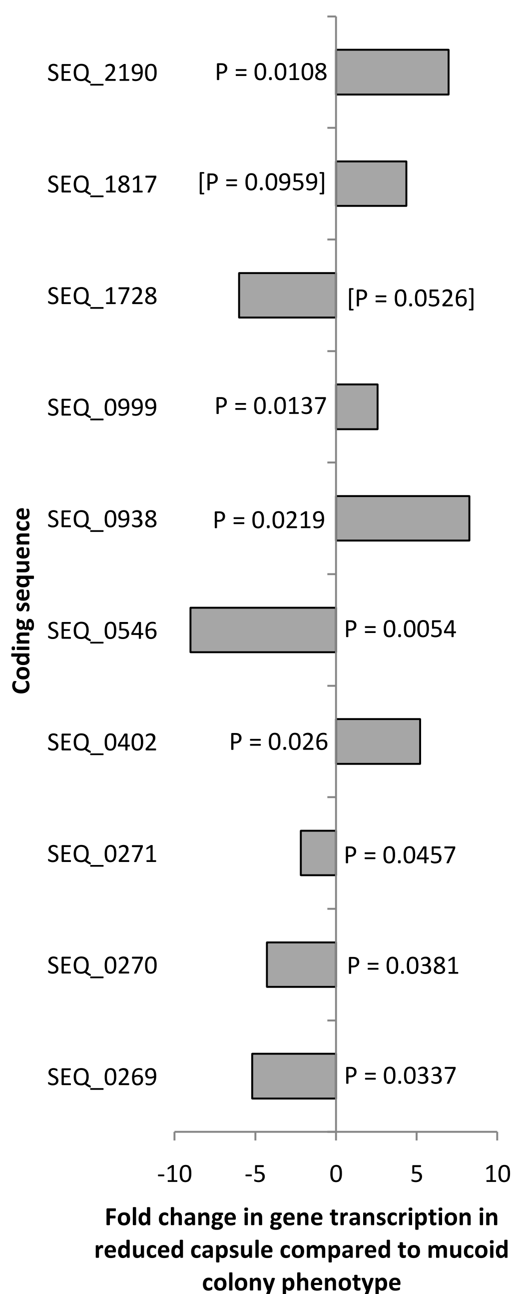


Nutrient acquisition
Capsule production
Surface anchored
Pathogenicity/adaptation/chaperone
Glycosylation
Amino acid biosynthesis
Secreted protein
Nutrient metabolism
Membrane protein
Regulator
Hypothetical
Laterally acquired elements
DNA modification
Pseudogene

436

Molecular BioSystems Accepted Manuscript

437 **Figure 3.** Graph showing the fold change in gene transcription from qPCR of reduced
438 capsule compared to mucoid colony phenotypes, where a positive value indicates up-
439 regulation in reduced capsule compared to mucoid colonies. P values indicate the
440 significance of differential transcription; square brackets indicate values that are not
441 significant.



442

443

Molecular BioSystems Accepted Manuscript

444 **Supplementary Material:**

445 **Table S1**. List of primers used in this study.

Primer name	Primer sequence (5' to 3')
SEQ0546 transcript qPCR F1	CTAGTGTAGCAGAAACAAC
SEQ0546 transcript qPCR R1	CAGCTGCACCATAATTG
SEQ0938 transcript qPCR F1	GTATGCTAGGAATGATG
SEQ0938 transcript qPCR R1	GCTCAGTTTGACTTAG
SEQ0999 transcript qPCR F1	CCATATCTTGCTCATC
SEQ0999 transcript qPCR R1	CAAACCACCAATGCAG
SEQ0269 transcript qPCR F1	GTGTTCAGCAGCAAACC
SEQ0269 transcript qPCR R1	CACATAGTCTTCAATGCG
SEQ0270 transcript qPCR F1	GCATATCGCCAATCAAG
SEQ0270 transcript qPCR R1	CATTGCTATAAGCGCTTG
SEQ0271 transcript qPCR F1	CATTCAATTCATCGTCGAG
SEQ0271 transcript qPCR R1	GAGCTCGAAGTTGGAGTC
SEQ1728 transcript qPCR F1	CAGATACGTACAATACAAATGATG
SEQ1728 transcript qPCR R1	CTTCTATTATCTAAAATTTCGTC
SEQ2190 transcript qPCR F1	GCGAAAGAGGAAGCTAAGAC
SEQ2190 transcript qPCR R1	CTTTATCAGCCGTAGCTTG
SEQ0402 transcript qPCR F1	CAGCGACTACCCTAGCAG
SEQ0402 transcript qPCR R1	CATTTTTAAGCTCGTTAGCG
SEQ1817 transcript qPCR F1	GCTCCTGCACCGAAAG
SEQ1817 transcript qPCR R1	CTAAAGAATGGATGGCTTGC
gyrA forward	CTATGAAGCGATGGTCCGTATGG
gyrA reverse	CCATAGAACCAAAGTTTCCATGACC

446

Molecular BioSystems Accepted Manuscript

447 **Table S2.** List of genes that were significantly differentially transcribed from transcriptome
448 sequencing between the reduced capsule and mucoid colony phenotypes of *Se*1691.

Gene ID	Description	Functional Group	Mean fold change in reduced capsule compared to mucoid	Q value
SEQ_0044	Aldehyde-alcohol dehydrogenase 2 [includes: alcohol dehydrogenase; acetaldehyde dehydrogenase] (*adhE*)	Nutrient metabolism	6.32	0.0493
SEQ_0090	Putative collagen-like surface-anchored protein	Surface anchored	3.76	0.0015
SEQ_0103	Putative membrane protein	Membrane protein	2.86	0.0298
SEQ_0203	Putative exported protein	Secreted protein	2.31	0.0482
SEQ_0225	Putative exported protein	Secreted protein	-8.10	0.0015
SEQ_0231	Putative Mga-like regulatory protein	Regulator	-3.05	0.0015
SEQ_0256	Putative cell surface-anchored protein	Surface anchored	-5.99	0.0015
SEQ_0257	Putative carbohydrate-binding exported protein	Nutrient acquisition	-2.42	0.0285
SEQ_0269	Hyaluronan synthase	Capsule production	-4.68	0.0015
SEQ_0270	UDP-glucose 6-dehydrogenase	Capsule production	-3.95	0.0015
SEQ_0271	UTP--glucose-1-phosphate uridylyltransferase 1	Capsule production	-2.21	0.0285
SEQ_0280	Putative collagen-like cell surface-anchored protein	Surface anchored	4.60	0.0433
SEQ_0281	Putative membrane protein	Membrane protein	2.39	0.0120
SEQ_0282	Putative glucitol/sorbitol-specific phosphotransferase system (PTS), IIA component	Nutrient acquisition	5.91	0.0290
SEQ_0315	Putative N-acetylmannosamine-6-phosphate 2-	Nutrient metabolism	-6.01	0.0466

Molecular BioSystems Accepted Manuscript

	epimerase			
SEQ_0316	Sugar phosphotransferase system (PTS), IIBC component	Nutrient acquisition	-3.87	0.0015
SEQ_0317	Putative glucose-specific phosphotransferase system (PTS), IIABC component	Nutrient acquisition	-4.27	0.0015
SEQ_0320	Putative N-acetylmannosamine-6-phosphate 2-epimerase (pseudogene)	Nutrient metabolism	-8.22	0.0015
SEQ_0321	Extracellular solute-binding lipoprotein	Nutrient acquisition	-11.03	0.0015
SEQ_0322	Putative transport system permease	Nutrient acquisition	-10.34	0.0015
SEQ_0323	Putative transport system permease	Nutrient acquisition	-9.18	0.0015
SEQ_0324	Conserved hypothetical protein	Nutrient acquisition	-8.44	0.0015
SEQ_0325	Putative membrane protein	Nutrient acquisition	-7.31	0.0015
SEQ_0326	Putative N-acetylneuraminate lyase	Nutrient metabolism	-7.54	0.0015
SEQ_0327	ROK family protein	Regulator	-6.65	0.0015
SEQ_0329	Putative transcription regulator	Regulator	-7.22	0.0015
SEQ_0350	MerR family regulatory protein	Regulator	-8.05	0.0015
SEQ_0351	Putative glutamine synthetase	Amino acid biosynthesis	-5.96	0.0015
SEQ_0352	Putative peptide binding protein	Nutrient acquisition	-6.08	0.0015
SEQ_0353	Di-tripeptide transport system permease protein	Nutrient acquisition	-3.98	0.0015
SEQ_0354	Di-tripeptide transporter permease protein	Nutrient acquisition	-3.90	0.0015
SEQ_0355	Di-tripeptide transporter ATP-binding protein	Nutrient acquisition	-3.89	0.0015
SEQ_0356	Di-tripeptide transporter ATP-binding protein	Nutrient acquisition	-3.98	0.0081
SEQ_0377	Putative protein F2 like fibronectin-binding protein (pseudogene)	Secreted protein	-3.35	0.0466

Molecular BioSystems Accepted Manuscript

	FNZ			
SEQ_0378	Putative regulatory protein-RofA related	Regulator	-4.32	0.0015
SEQ_0382	Hypothetical protein	Hypothetical	2.28	0.0408
SEQ_0402	Putative cell surface-anchored protein	Surface anchored	3.94	0.0055
SEQ_0430	Sucrose-6-phosphate hydrolase	Nutrient metabolism	2.19	0.0493
SEQ_0490	Putative membrane protein	Membrane protein	-5.63	0.0015
SEQ_0504	Putative exported protein	Secreted protein	-4.93	0.0015
SEQ_0546	Streptolysin S precursor	Pathogenicity/adaptation/chaperones	-4.60	0.0015
SEQ_0671	Probable potassium transport system protein	Nutrient acquisition	-3.59	0.0015
SEQ_0696	DNA/RNA non-specific endonuclease	Pathogenicity/adaptation/chaperones	-2.21	0.0224
SEQ_0721	Immunoglobulin G-binding protein G precursor (IgG-binding protein G) LPXTG	Surface anchored	-2.35	0.0179
SEQ_0817	Hypothetical phage protein	Phage	-2.14	0.0168
SEQ_0855	Putative collagen-like surface-anchored protein	Surface anchored	4.83	0.0234
SEQ_0862	GntR family regulatory protein	Regulator	-4.22	0.0015
SEQ_0866	Beta-galactosidase 3	Pseudogene	-3.83	0.0042
SEQ_0896	Putative exported protein	Secreted protein	3.43	0.0015
SEQ_0931	Putative permease	Nutrient acquisition	-9.03	0.0015
SEQ_0932	ABC transporter ATP-binding protein	Nutrient acquisition	-8.97	0.0015
SEQ_0938	Immunoglobulin G-endopeptidase (*ideS*) / Mac/ Secreted immunoglobulin binding protein (Sib38)	Pathogenicity/adaptation/chaperones	4.62	0.0015
SEQ_0947	Conserved hypothetical protein	Hypothetical	16.56	0.0015
SEQ_0948	Putative membrane protein	Membrane protein	6.46	0.0015
SEQ_0949	Putative exported protein	Secreted protein	15.15	0.0015
SEQ_0950	Conserved hypothetical protein	Hypothetical	7.80	0.0015
SEQ_0951	Glycosyl transferase	Glycosylation	2.93	0.0015

Molecular BioSystems Accepted Manuscript

	family protein			
SEQ_0999	*ideE*; IgG endopeptidase	Pathogenicity/adaptation/chaperones	2.88	0.0015
SEQ_1001	Putative exported protein	Secreted protein	3.63	0.0015
SEQ_1278	Putative surface-anchored 5'-nucleotidase	Surface anchored	-3.66	0.0015
SEQ_1281	ABC transporter, ATP-binding/permease protein (pseudogene)	Pseudogene	-3.64	0.0043
SEQ_1283	ABC transporter, ATP-binding/permease protein	Nutrient acquisition	-3.55	0.0029
SEQ_1291	Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase family protein	Nutrient acquisition	-23.48	0.0015
SEQ_1328	Hypothetical protein (pseudogene)	Pseudogene	2.46	0.0015
SEQ_1345	Transport system membrane protein	Nutrient acquisition	2.26	0.0068
SEQ_1346	Extracellular solute-binding protein	Nutrient acquisition	2.40	0.0015
SEQ_1351	Conserved hypothetical protein	Hypothetical	-6.75	0.0015
SEQ_1352	LysR family regulatory protein	Regulator	-2.34	0.0029
SEQ_1375	Sugar phosphotransferase system (PTS), mannose/fructose family, IIA component	Nutrient acquisition	2.67	0.0055
SEQ_1505	Putative exported protein	Secreted protein	6.32	0.0015
SEQ_1543	Putative exported protein	Secreted protein	3.80	0.0015
SEQ_1548	Putative alkaline phosphatase synthesis sensor protein	Regulator	-3.40	0.0015
SEQ_1549	Alkaline phosphatase synthesis transcriptional regulatory protein	Regulator	-3.46	0.0343
SEQ_1654	Conserved hypothetical protein	Hypothetical	-3.31	0.0015
SEQ_1661	Metal-dependent transcriptional regulator	Regulator	-2.11	0.0336
SEQ_1728	Exotoxin L precursor, *seeL*	Pathogenicity/adaptation/chaperones	-4.71	0.0015

Molecular BioSystems Accepted Manuscript

SEQ_1768	Putative cysteine synthase	Amino acid biosynthesis	2.45	0.0015
SEQ_1800	Putative exported protein	Secreted protein	-8.16	0.0015
SEQ_1817	Collagen-like surface-anchored protein	Surface anchored	4.15	0.0015
SEQ_1821	Putative lipoprotein	Nutrient acquisition	-13.75	0.0015
SEQ_1903	Extracellular solute-binding protein	Nutrient acquisition	-2.15	0.0427
SEQ_1941	Putative membrane protein	Membrane protein	3.80	0.0015
SEQ_1949	Putative secreted alpha-amylase	Nutrient acquisition	2.31	0.0317
SEQ_1957	Streptococcal histidine triad protein	Nutrient acquisition	-6.04	0.0015
SEQ_1958	Laminin binding protein	Nutrient acquisition	-8.05	0.0015
SEQ_1959	Putative cell surface-anchored protein	Surface anchored	3.67	0.0015
SEQ_1960	Conserved hypothetical protein	Hypothetical	-4.55	0.0015
SEQ_1977	Type I restriction-modification system M protein	DNA modification	-2.07	0.0290
SEQ_2010	Thioredoxin (pseudogene)	Pseudogene	-2.95	0.0015
SEQ_2014	Streptokinase	Pathogenicity/adaptation/chaperones	-3.52	0.0015
SEQ_2125	ABC transporter ATP-binding protein	Nutrient acquisition	-2.65	0.0015
SEQ_2126	Transport system permease protein	Nutrient acquisition	-2.33	0.0215
SEQ_2127	Putative lipoprotein	Nutrient acquisition	-2.34	0.0144
SEQ_2128	Putative lipoprotein	Nutrient acquisition	-2.24	0.0267
SEQ_2180	Putative cell surface-anchored protein	Surface anchored	4.24	0.0015
SEQ_2190	Putative cell surface-anchored protein	Surface anchored	4.75	0.0015
SEQ_2196	PadR family regulatory protein	Regulator	-8.31	0.0015
SEQ_2197	Hypothetical protein	Hypothetical	-6.89	0.0015
SEQ_2198	Hypothetical protein	Hypothetical	-5.35	0.0015
peg 1573	Mobile element protein	Phage	3.62	0.0015
peg 1265	Hypothetical protein	Hypothetical	5.67	0.0015
peg 1583	N-acetylmannosamine-6-phosphate 2-epimerase	Nutrient metabolism	-6.70	0.0029

Molecular BioSystems Accepted Manuscript

peg 1601	N-acetylmannosamine-6-phosphate 2-epimerase	Nutrient metabolism	-17.44	0.0053
peg 1647	Hypothetical protein	Hypothetical	-7.17	0.0015
peg 1771	Hypothetical protein	Hypothetical	-3.91	0.0229

449

Molecular BioSystems Accepted Manuscript