

1 A high-throughput genomic screen identifies a role for the plasmid- 2 borne Type II secretion system of *Escherichia coli* O157:H7 (Sakai) 3 in plant-microbe interactions 4

5 Ashleigh Holmes¹, Leighton Pritchard^{1,2}, Peter Hedley¹, Jenny Morris¹, Sean P. McAteer³, David L.
6 Gally³ and Nicola J. Holden^{1*}

7 ¹Cellular and Molecular Sciences, James Hutton Institute, Dundee, DD2 5DA, UK.

8 ²Strathclyde Institute for Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, G4
9 ORE, UK

10 ³The Roslin Institute Division of Infection and Immunity, University of Edinburgh, R(D)SVS, The Roslin
11 Institute Building, Easter Bush, EH25 9RG, UK.

12 * corresponding author: Nicola J. Holden, nicola.holden@hutton.ac.uk

13
14 Keywords: STEC, VTEC, general secretory pathway, pO157-cured, spinach, enrichment screen

15
16 Word count: 6200 words

17 Abstract

18 Food-borne illness arising from Shiga-toxigenic *Escherichia coli* (STEC) is often linked to consumption
19 of fruit and vegetables as the bacteria have the ability to interact with plants and use them as
20 alternative or secondary hosts. The initial stages of the interaction involve chemotaxis, attachment
21 and potentially, responding to the early stages of microbe perception by the plant host. We used a
22 high-throughput positive-selection approach to identify early interaction factors of *E. coli* O157:H7
23 isolate Sakai to spinach. A bacterial artificial chromosome (BAC) clone library was quantified by
24 microarray hybridisation, and gene loci enrichment measured using a Bayesian hierarchical model.
25 The screen of four successive rounds of short-term (2 hour) interaction with spinach roots produced
26 in 115 CDS credible candidates, comprising seven contiguous genomic regions. Two candidate
27 regions were selected for functional assessment: a chaperone-usher fimbrial gene cluster (*loc6*) and
28 the pO157 plasmid-encoded type two secretion system (T2SS). Interaction of bacteria with spinach
29 tissue was reduced in the absence of the pO157 plasmid, which was appeared to involve the T2SS
30 EtpD secretin protein, whereas loss of *loc6* did not impact interactions. The T2SS genes, *etpD* and
31 *etpC*, were expressed at a plant-relevant temperature of 18 °C, and *etpD* expressed *in planta* by *E.*
32 *coli* Sakai on spinach plants. Thus, a whole genome screening approach using a combination of
33 computational modelling and functional assays has identified a novel function for STEC T2SS in
34 interactions with plant tissue.

35 1. Introduction

36 Shiga-toxigenic *Escherichia coli* (STEC) (or verocytotoxigenic *E. coli*, VTEC) including the predominant
37 serotype O157:H7, are significant zoonotic and food-borne pathogens, across the globe. Although
38 ruminant farm animals are the primary reservoir for STEC, they can be transmitted through the food-
39 chain on edible plants and plant-derived foodstuffs account for a large proportion (>50%) of food-
40 borne illness in the USA [1]. However, animals remain the primary source of STEC on plants, either
41 through direct application of manure/biosolids as fertilisers, or more likely via contaminated
42 irrigation water [2].

43 STEC has been shown to interact with plants and can colonise them as secondary hosts [3].
44 Colonisation of STEC has been demonstrated on plant roots and in the rhizosphere [4-6], a
45 favourable environment for bacteria that is rich in root exudates, which include a source of nutrients
46 [7] and chemoattractants [8]. Numbers of *E. coli* recovered from roots often are greater than that
47 from the leaves [6] and STEC has been shown to persist in soil and on plants for extended periods,
48 e.g. >75 days [9].

49 Initial interactions in host colonisation involve chemotaxis, adherence and response to host
50 perception. Since attachment is considered a prerequisite for successful colonisation, various
51 approaches have been taken to identify adherence factors. The genome of STEC serotype O157:H7
52 isolate Sakai [10] encodes up to 14 fimbriae gene loci. Many of the *E. coli* adhesins show specificity in
53 their host interactions, conferring a degree of tissue tropism for different *E. coli* pathotypes [11].
54 Curli, long polar fimbriae (Lpf), *Escherichia coli* common pilus (ECP), flagella and the T3SS have all
55 been implicated in plant associated adherence of STEC [12-17], but several others STEC adherence
56 gene clusters have yet to be functionally characterised. As such, we hypothesised that the STEC
57 genome encodes additional uncharacterised factors that facilitate initial interactions with plant
58 tissue. To identify which STEC genomic regions confer an advantage to colonisation of plant roots, a
59 positive-selection screening approach was taken using an *E. coli* Sakai BAC clone library for short-
60 term (2 hours) interactions with plant roots. Spinach was selected as it is relevant to large-scale STEC
61 outbreaks [18], and we have previously shown specific adherence to spinach roots [14, 15, 6]. High-
62 throughput screening enables wholesale analysis and previous global transcriptomic analysis has
63 shown induction of STEC fimbrial and afimbrial adhesins in lettuce leaf lysates [19-21]. In a similar
64 manner, high-throughput negative- and positive-selection approaches have identified colonisation
65 factors, e.g. a random mutant library of *Pseudomonas fluorescens* was used to identify plant
66 colonisation factors [22], and signature tagged mutagenesis and a bacterial artificial chromosome
67 (BAC) library have been used to investigate STEC interactions with bovine mucus [23, 24]. Therefore,
68 we used a BAC clone library of *E. coli* O157:H7 isolate Sakai that was previously used to identify

69 genetic loci that enhanced adherence to bovine epithelial cells, and promoted bacterial growth in
70 bovine mucus [23]. *E. coli* Sakai was used because it was derived from a large outbreak arising from
71 contamination of white radish sprouts [25]. The approach involved whole-genome interrogation
72 using microarrays and Bayesian analysis to compare the library clones prior- and post- spinach root
73 inoculation.

74 The BAC library screen identified several contiguous *E. coli* Sakai chromosomal and plasmid regions
75 that enriched following interaction with spinach roots, present on *E. coli* O157:H7-specific genomic
76 segments known as S-loops [26]. Candidate regions that included annotated adherence factors were
77 taken forward for characterisation. Functional analysis identified the plasmid-borne Type II Secretion
78 System (T2SS) as a factor that conferred increased adherence for *E. coli* O157:H7 Sakai to both
79 spinach roots and leaves.

80 **2. Materials and Methods**

81 **2.1 Bacterial strains and media**

82 *E. coli* O157:H7 isolate Sakai, hereafter *E. coli* Sakai [10] and its derivatives were grown in either
83 lysogeny broth (LB) or MOPS medium [27] supplemented with 0.2 % glucose (or glycerol where
84 indicated), 10 μ M thiamine and MEM essential and non-essential amino acids (Sigma M5550 and
85 M7145) termed rich defined MOPS (RD-MOPS) media. Antibiotics were included where necessary to
86 maintain transformed plasmids at the following concentrations: 50 μ g/ml kanamycin (Kan), 25 μ g/ml
87 chloramphenicol (Cam), 10 μ g/ml Tetracycline (Tet), 50 μ g/ml ampicillin (Amp).

88 **2.2 Plant propagation**

89 Spinach (*Spinacia oleracea*) cultivar Amazon seeds (Sutton Seeds, UK) were grown in hydroponics for
90 the BAC screen. Seeds were germinated on distilled water agar (0.5 % w/v) and after 3-5 days
91 transplanted into pots containing autoclaved vermiculite and sterile 0.5 x Murashige and Skoog (MS)
92 medium (Sigma Aldrich, USA) with no carbon supplement. Plants were maintained under
93 environmental cabinet conditions as above for 4-6 weeks. Spinach was grown similarly for BAC clone
94 adherence assays and confocal microscopy of roots, for hydroponics plants in sterile hydroponic tubs
95 (Greiner, UK) containing perlite instead of vermiculite (optimal for microscopy of roots). Spinach
96 was grown in compost for adherence assays and confocal microscopy of leaves. Seedlings were
97 grown in an environmental cabinet with a light intensity of 150 μ mol m² s⁻¹ (16 hour photoperiod)
98 for a further 21 days at 22 °C. Compost-grown plants were germinated and maintained in individual
99 posts with commercial compost and under glasshouse conditions 22 °C (16 h of light, 8 h of dark)
100 with 130–150 μ mol m² s⁻¹ light intensity and 40% humidity.

101 2.3 Bacterial Artificial Chromosome Library screen for adherence to spinach 102 roots

103 The BAC library contained a partial *Hind*III digest of *E. coli* Sakai genome cloned into pV41 vector,
104 and together with the spinach root adherence approach, is described in detail in (Accompanying DiB
105 paper DIB-S-20-00975).

106 2.4 Microarray hybridisation and data analysis

107 The microarray chip used for the analysis, a 8 x 15k *E. coli* gene expression array, *E. coli* v.2 (Agilent
108 product number G4813A-020097) and gDNA extraction is described in detail in (*Submitted dataset to*
109 *DiB*). Gene enrichment data is deposited with ArrayExpress with accession numbers for the
110 adherence treatment: E-MTAB-5923 and control treatment: E-MTAB-5924. A complete description
111 of the data analysis is provided at https://widdowquinn.github.io/SI_Holmes_etal_2017/
112 (doi:10.5281/zenodo.822825) but briefly, probe intensity data was subjected to QA and clean-up in
113 which three problematic probes in a single treatment arm replicate were replaced with values
114 interpolated from the other two treatment replicates. Array intensities were quantile normalised
115 separately for control and treatment arms, and each probe annotated by BLASTN match to the most
116 recent CDS annotations for the *E. coli* DH10B and Sakai isolates (NCBI accessions:
117 GCF_000019425.1_ASM1942v1, GCF_000008865.1_ASM886v1). Only probes that unambiguously
118 matched to a single Sakai or DH10B CDS were taken forward in the analysis (8312 unique probes,
119 6084 unique CDS, 49872 datapoints).

120 A Bayesian hierarchical model was fit to the array intensity data. This model treats growth and
121 amplification ('control' and 'treatment' arms) and adherence to roots ('treatment arm only') as
122 additive linear effects describing the relationship between the measured intensity for each probe i
123 before (x_i) and after (y_i) each replicate experiment. In this model, parameters for the linear
124 components were pooled either by the CDS from which the probes are derived (for gradients: β and
125 δ , with corresponding index for the associated CDS $j[i]$), or the array used for that replicate (for
126 offsets: α and γ , with corresponding index for the array/replicate $k[i]$). A binary 1/0 value (t_i) was
127 used to indicate whether a specific experiment did or did not include the spinach root adherence:

$$\hat{y}_i = \alpha_{k[i]} + \beta_{j[i]}x_i + \gamma_{k[i]}t_i + \delta_{j[i]}t_ix_i$$

$$y_i \sim N(\hat{y}_i, \sigma_y^2); \sigma_y \sim U(0, \infty)$$

$$\alpha_{k[i]} \sim \text{Cauchy}(\mu_\alpha, \sigma_\alpha^2); \sigma_\alpha \sim U(0, 100)$$

$$\beta_{j[i]} \sim \text{Cauchy}(\mu_\beta, \sigma_\beta^2); \sigma_\beta \sim U(0, 100)$$

$$\gamma_{k[i]} \sim \text{Cauchy}(\mu_\gamma, \sigma_\gamma^2); \sigma_\gamma \sim U(0, 100)$$

$$\delta_{j[i]} \sim \text{Cauchy}(\mu_{\delta}, \sigma_{\delta}^2); \sigma_{\delta} \sim U(0, 100)$$

128 The model was fit using PyStan 2.12.0.0 under Python 3.6, with two chains each of 1000 iterations,
129 to estimate parameter values: $\alpha_{k[i]}$ - the array-level offset due to growth for each replicate; $\beta_{j[i]}$ - the
130 CDS-level influence of the growth step on probe intensity; $\gamma_{k[i]}$ - the array-level offset due to
131 treatment/passage for each replicate; $\delta_{j[i]}$ - the CDS-level influence of treatment/passage on probe
132 intensity; $\mu_{\alpha}, \mu_{\beta}, \mu_{\gamma}, \mu_{\delta}$ - the pooled distribution means for each of the four main equation
133 parameters; $\sigma_{\alpha}, \sigma_{\beta}, \sigma_{\gamma}, \sigma_{\delta}$ - the scale values for the pooled distributions for each of the four main
134 equation parameters; and σ_y - the variance due to irreducible measurement error.

135 The CDS with index $j[i]$ was considered to be associated with an advantageous effect on adherence
136 (positive selection pressure) if the median estimated value of $\delta_{j[i]}$ was positive, and the
137 corresponding 50% credibility interval did not include zero. A similar interpretation was used to infer
138 an advantageous effect on *in vitro* growth/amplification from estimates of $\beta_{j[i]}$. Goodness of the
139 model fit was estimated using 10-fold crossvalidation. The model is described in full in an interactive
140 Jupyter notebook in Supplementary Information.

141 2.5 Molecular methods

142 All primers and plasmids are listed in Table S2. To identify BAC clones containing the *etp* operon,
143 bacterial pools consisting of 48 clones of the library were screened by PCR for *etpD* and *etpO* genes
144 using primers *etpD*.RT.F, *etpD*.RT.R, *etpO*.F, *etpO*.R. Individual clones in the pool were then screened
145 using the same primers, identifying clone BAC2B5. BAC2B5 sequence was determined from primer
146 walking near HindIII sites in pO157 with primers specific to the pVG1 vector. PCR products amplified
147 using primer pairs BAC2B24F and pVG1; BAC2B5F and pVG1.R were Sanger sequenced. This
148 confirmed the sequence from the BAC vector pVG1 to the upstream and downstream sequence at
149 pO157 HindIII 87463. *E. coli* strain Sakai was cured of the pO157 plasmid by plasmid incompatibility
150 as described by [28]. In short, Sakai was transformed with pBeloBAC11 which has the same
151 incompatibility as pO157. Transformants were subcultured three times in LB+Cam to cure the
152 pO157. Plasmid curing was confirmed by PCR for *tox*B, *hly*AB and *etpO*. The pBeloBAC11 was cured
153 by sub-culturing three times in LB without selection. Loss of pBeloBAC11 was confirmed by loss of
154 Cam resistance and by PCR for the vector using primers T7 promoter and Cml_rev. The pO157-cured
155 and WT strains were whole genome sequenced from a paired-end library to generate short-read
156 (Illumina) sequences (ENA accessible number: ERS4383229 – accessible 30-Jun-2020), which were
157 annotated using PROKKA [29] for Blastp [30] comparisons, using the reference Sakai sequence
158 (BA000007.3) on the Galaxy platform [31]. A defined deletion in the *E. coli* O157:H7 isolate Sakai
159 *etpD* gene (pO157p03) and *loc6* fimbrial locus (ECs1276-1280) was constructed using allelic

160 exchange as previously described [32, 14] using constructed vectors pAH005 (*loc6*) and pAH006
161 (*etpD*), respectively. The Sakai Δ *etpD* strain was cured for resistance to tetracycline by transforming
162 the mutant with FLP recombinase expressing plasmid pCP20 [33]. Deletions were confirmed by PCR
163 and Sanger sequencing, and for the Sakai Δ *etpD* strain by whole genome sequencing and BLASTn
164 analysis to confirm loss of the CDS for pO157p03 locus. The promoterless *etpD* gene was PCR
165 amplified (primers EtpD.Xba.pSE and EtpD.Hind.pSE) and cloned into the IPTG inducible plasmid
166 pSE380 to create pAH007 and complement the mutation *in trans*. For the GFP transcriptional
167 reporters, the 5'UTR of *etpC* and *etpD* was PCR amplified (primers EtpC.XbaI.F, EtpC.XbaI.R,
168 pKC_EtpD.XbaF, pKC_EtpD.XbaR) and cloned into pKC026 using XbaI, creating the transcriptional
169 fusions pAH008 (*etpC*) and pAH009 (*etpD*), respectively.

170 2.6 Bacterial adhesion assays on plant tissues

171 Adherence assays were performed as described in [15]. In short, plant tissues were washed and
172 incubated in bacterial suspension ($\sim 1 \times 10^7$ cfu/ml in sterile PBS; OD₆₀₀ = 0.02) statically for two hours
173 at 18 °C. Plant samples were vigorously washed 3 times in sterile PBS by mixing on a vortexer,
174 weighed then homogenised with a sterile pestle and mortar. Samples were serially diluted and
175 plated on MacConkey's agar with appropriate antibiotics for bacterial counts. Measurements of *E.*
176 *coli* Sakai wild type and *etpD* knockout, and Sakai Δ *etpD* transformed with the empty vector
177 (pSE380) and *etpD* complement (pSE-*etpD*), were performed separately in batches of five biological
178 replicates on independent leaf or root tissues as appropriate. Four batches were obtained for leaf
179 tissue, and six for root tissue.

180 The bacterial recovery data (logCFU) was fit to a linear model describing additive non-interacting
181 effects due to: *E. coli* Sakai adhesion (α); the modification of wild-type adhesion due to knockout of
182 *etpD* (β); the introduction of empty pSE380 plasmid into the knockout background (γ); the effect of
183 introducing pSE-*etpD* with respect to introduction of the empty vector in the knockout background
184 (δ); and batch effects ($\phi_{1..n}$). The data were fit using PyStan 2.16.0.0 under Python 3.6, and the
185 parameter estimates for β and δ and their 50% and 95% credibility intervals were used to infer the
186 effects of knockout and complementation of *etpD*, respectively. These estimates represent the
187 change in recovered bacterial counts as a result of the specific modification (loss or gain of *etpD*)
188 with respect to the appropriate control. The model fit is described in full in a Jupyter notebook
189 (https://widdowquinn.github.io/SI_Holmes_etal_2017/notebooks/04-etpD.html).

190 2.7 Bacterial adhesion to abiotic surfaces

191 Bacterial strains were cultured in LB at 37°C, 200rpm, for 16 hours then washed in fresh LB, RD
192 MOPS glucose or RD MOPS glycerol. To assess initial attachment, the OD₆₀₀ was adjusted to 0.5 for
193 2 hours incubation in the microtiter plate; for early biofilm formation, the OD₆₀₀ was adjusted to 0.02

194 for 24 hours incubation. 200 μ l was aliquoted in quadruplicate in an untreated 96 well plate (VWR,
195 UK). The plate was incubated at 18°C statically before measuring adherent bacteria by Crystal Violet
196 as described in [34].

197 2.8 Analysis of bacterial fluorescence *in vitro*

198 Gene expression was measured from *E. coli* Sakai transformed with pAH008 or pAH009 following
199 growth for ~18 hours in LB medium + Chl at 37 °C, 200 rpm before diluting 1:100 into 15 ml RD
200 MOPS medium supplemented with 0.2 % glucose or glycerol. Cultures were incubated statically at 18
201 °C and samples periodically removed and measured for cell density and GFP fluorescence. GFP
202 fluorescence was measured in triplicate 200 μ l volumes in a 96 well plate using GloMax plate reader
203 (Promega). *E. coli* Sakai transformed with the vector control plasmid pKCO26 was included as a
204 control for background fluorescence. Fluorescence was plotted against OD₆₀₀ and a quadratic line of
205 best fit obtained. This was used to correct readings for background fluorescence. Corrected data was
206 normalised to cell density (OD₆₀₀) and values plotted using GraphPad Prism software for two
207 experimental repeats.

208 2.9 Confocal microscopy

209 Fully expanded 4-week-old spinach leaves were infiltrated, by pressure injection using a 1 ml
210 needleless syringe into the abaxial epidermis, with approx. 10⁶ cfu *E. coli* Sakai + pAH009 + *pmKate*
211 and the plants maintained in an environmental cabinet until observed four days later. Two leaves on
212 two individual plants were infiltrated per experiment and the experiment repeated on spinach
213 plants propagated several weeks later. High inoculum levels ensured sufficient cells for observation
214 since we have previously shown that *E. coli* Sakai is unable to proliferate in the apoplast of spinach
215 and remains in a persistent state [17]. Leaf segments were infiltrated with sterile distilled water, to
216 displace air from the apoplastic spaces between the spongy mesophyll cells, prior to mounting
217 abaxial side up on microscope slides using double-sided tape. For spinach roots, 5 weekold spinach
218 were grown under hydroponic culture as described, the 0.5x MS was removed and replaced with
219 10ml 0.5x MS inoculated with 10⁸ cfu bacteria. After four days in environmental cabinet conditions,
220 the tub was flooded with sterile PBS to displace the perlite from the roots as non-invasively as
221 possible. The leafy part of the plant was removed from the root by sterile scalpel cutting
222 approximately 5mm below the cotyledon. After a further two washes in PBS, the root was mounted
223 on a microscope slide, flooded with sterile PBS, and the coverslip held in place with double-sided
224 tape.

225 Mounted plant tissue samples were observed using a Nikon A1R confocal laser scanning microscope
226 mounted on a NiE upright microscope fitted with an NIR Apo 40x 0.8W water dipping lens and GaAsP
227 detectors. Images represent false-coloured maximum intensity projections as indicated, produced

228 using NIS-elements AR software. GFP (green) and chlorophyll (blue) were excited at 488 nm with the
229 emissions at 500-530 nm and 663-737 nm respectively, and mKate (RFP) was excited at 561 nm with
230 emission at 570-620 nm (magenta).

231 **3 Results**

232 **3.1 Interaction screen using an *E. coli* isolate Sakai BAC clone library**

233 To identify candidate gene loci for *E. coli* O157:H7 isolate Sakai (hereafter: *E. coli* Sakai) that
234 conferred an advantage to spinach root tissue interactions, a Sakai BAC clone library was employed
235 hosted in *E. coli* strain DH10B, which is derived from a K-12 strain and in our hands is a poor
236 coloniser of plants [6]. A differential screen compared BAC clones inoculated with spinach roots to
237 BAC clones treated similarly but in the absence of spinach roots. The BAC library was inoculated with
238 freshly harvested spinach roots for two hours (insufficient time for bacterial proliferation) in four
239 successive rounds to enrich for interactions. Loosely-attached and non-adherent bacteria were
240 excluded between each round, so that the only strongly-adherent population were used for
241 subsequent inoculation rounds, since these are most likely to be retained as 'successful colonisers'.
242 Each round resulted in successive reductions of the number of bacteria recovered from the roots as
243 selectivity increased, with a 400-fold reduction between round 1 and 2 from 6×10^5 cfu/ml to $1.6 \times$
244 10^3 cfu/ml, which necessitated an amplification step after the second round to ensure that there
245 were sufficient bacteria for subsequent selection rounds 3 and 4. An additional amplification step
246 after round 4 ensured sufficient gDNA for hybridisation to the microarray. The no-plant negative
247 control treatment did not include spinach root tissue, where the bacteria were inoculated into
248 medium and suspended in PBS alone, to account for gene loci in the BAC clone library that may have
249 conferred an advantage during the amplification steps between round 2 & 3 and after round 4. After
250 four rounds of selection and enrichment, a total of 7.17×10^8 cfu/ml of bacteria were recovered
251 from the plant-treatment compared to 1.13×10^9 cfu/ml of bacteria from the negative control
252 treatment and taken forward for gene abundance analysis.

253 Gene abundance in pools of BAC clone gDNA was quantified on a DNA microarray before (i.e. input
254 pools) and after selection (output pools), for both plant and no-plant treatments (dataset submitted
255 to DiB DIB-S-20-00975). A Bayesian hierarchical model was fitted to the probe intensity data to
256 estimate for each CDS in the *E. coli* DH10B and Sakai genomes a parameter representing the
257 selection pressure due to inoculation on the plant. A CDS was considered to be under positive
258 selective pressure (i.e. enriched) if its estimated value of this parameter was positive, and its 50%
259 credibility interval did not include zero. This resulted in 115 CDS with a credible positive effect on
260 adherence (Table S1).

261 3.2 Spinach root interactions enrich *E. coli* Sakai genes in six genomic regions 262 (S-loops)

263 The 115 CDS that correlated with adherence to spinach tissue comprised seven contiguous regions
264 of interest, of which 68 CDSs had existing functional annotation and 47 were annotated as
265 hypothetical proteins (Table S1). Enriched genes were grouped by chromosome / plasmid location
266 [10] and described in the context of the *E. coli* Sakai-specific S-loop designation [26]: S-loop 71; S-
267 loop 72 / prophage SpLE1; S-loops 85 / prophage Sp9; S-loop 225; S-loop 231; and pO157 (Fig. 1).

268 S-loop 71: a contiguous region in S-loop 71 was identified spanning 28 loci from ECs1272-ECs1296.
269 This region is equivalent to the genomic island OI#47 in STEC isolate EDL933, which is conserved in
270 STEC O157 serotypes [35], and includes the *loc6* fimbrial cluster, putative
271 hemagglutinin/haemolysin-like proteins and fatty-acid synthesis genes.

272 S-loop 72: Sakai prophage like element 1 (SpLE1) in S-loop 72 encodes 111 open reading frames
273 (ECs1299-ECs1409 [36]), of which 36 were enriched in interaction with spinach tissue, which we
274 termed SpLE1 (partial). Enriched genes included those for urea degradation *ureA,B,EFG*, of which
275 urease genes ECs1321-1327 were repressed in response to spinach root exudates [20]. Adhesion Iha
276 and AidA, encoded by ECs1360 and ECs1396 respectively, are also present in SpLE1, but were not
277 enriched in a contiguous region of 50 genes (ECs1349-1398).

278 S-loop 85: Prophage Sp9 in S-loop 85 includes a number of genes encoding non-LEE encoded (Nle)
279 effectors (*nleA*, *nleH2*, *espO1-2* and *nleG* [37]). This region was enriched in a separate study
280 investigating adherence to bovine primary tissue [23], and induction of *nleA* was induced in STEC
281 (EDL933) in response to lettuce leaf lysates [19].

282 S-loop 225: Gene loci in S-loop 225 (ECs4325 – 4341) are associated with fatty acid biosynthesis and
283 ECs4331 is annotated as a putative surfactin [26]. ECs4325-4340 were also induced in *E. coli* Sakai in
284 the presence of spinach leaf lysates [20].

285 S-loop 231: Gene loci in S-loop 231 (ECs4379 – 4387) are associated with heme utilisation and
286 transport and ECs4379 encodes a *chuS* heme oxygenase [38]. ECs4383/86/87 were induced in the
287 presence of spinach root exudates [20] and locus Z4912 (ECs4381) was induced for STEC isolate
288 EDL933 attached to radish sprouts [39].

289 pO157: pO157 p3,5,6, and 8 encode genes in the operon for a Type 2 secretion (T2SS) system. The
290 T2SS of STEC has been reported to play a role in adherence to mammalian host tissues [40]. The
291 pO157 has a role in biofilm formation, since a plasmid cured strain of *E. coli* Sakai was shown to have
292 reduced EPS production and did not generate hyperadherent variants (Lim et al., 2010).

293 Furthermore, the T2SS is an important virulence factor in many phytopathogens required for the
294 secretion of plant wall degrading enzymes (reviewed in [41]).

295 Analysis of the unclassified group (hypothetical genes) by InterProScan did not indicate any potential
296 roles in adherence and none were selected for functional analysis: 18 had no predicted functional
297 domains and six genes had a predicted transposase function (ECs1337-1340, Ecs3868-3869). Nine
298 were included above: four *nle* effectors in prophage Sp9; urease gene ECs1321; fatty acid synthesis
299 genes ECs4333 and 4335; and p79 and p81 from lipid operon *ecf*. Another four have domains of
300 unknown function (DUF).

301 On basis of gene annotation and any reference in the published literature, we focused on two
302 candidates that may have a function in adherence, as a key aspect of initial colonisation interactions:
303 the *loc6* gene cluster from S-Loop 71 since fimbriae are well described adherence factors, and the
304 T2SS genes on pO157, which are associated with biofilm formation. Therefore, the functional activity
305 of *loc6* and the pO157-based T2SS was assessed with spinach tissue using a series of deletion
306 mutants.

307 3.3 Functional characterisation of *loc6* fimbrial locus

308 A defined *loc6* (ECs1276-1280) deletion mutant was constructed in *E. coli* Sakai and its ability to
309 interact with spinach roots compared to the WT parental strain. There was no difference between
310 the numbers of the *Loc6* fimbriae-deficient bacteria recovered compared to wild-type, following a
311 two-hour incubation on spinach roots (Fig. 2). This suggested that the *loc6* fimbrial locus did not
312 confer a direct advantage on spinach roots, and it is possible that genes elsewhere in the contiguous
313 region were responsible for enrichment of the BAC clones (Table S1).

314 3.4 Functional characterisation of the pO157-encoded Type II secretion 315 system

316 3.4.1 A role for pO157 in spinach interactions

317 Candidate BAC clones containing TS22 genes (in *E. coli* DH10B background) were tested for their
318 ability to interact with spinach root tissue compared to the empty BAC vector, pV41 (also
319 transformed in DH10B). Clone BAC2B5, which encompasses the entire pO157 sequence, increased
320 adherence to spinach roots significantly ($p < 0.05$; students t test) compared to the pV41 vector-only
321 control (Fig. 3A). A plant-dependent specificity of the pO157 BAC2B5 clone was determined by
322 testing adherence to two non-plant surfaces. There was no significant difference in binding for clone
323 BAC2B5 compared to the vector-only control on natural wool (a biotic surface mimicking root
324 structures) (Fig. 3A; $p=0.9864$) or polystyrene (abiotic surface) (Crystal Violet (OD_{590nm}) mean of
325 BAC2B5: 0.0178 ± 0.0227 ; pVG1: 0.0236 ± 0.0303).

326 A role for the pO157 plasmid in interactions with spinach was confirmed by removal of the pO157
327 plasmid from *E. coli* Sakai. Plasmid loss was confirmed by PCR for the pO157 specific genes *tox*B,
328 *ehxA* and *etpO*, and from comparison of the whole-genome sequence and its isogenic parent (*E. coli*
329 Sakai WT). All the annotated pO157 plasmid coding sequences were absent in the pO157-cured
330 isolate except for two CDS associated with an IS element (IS629), while 100 % of the annotated
331 chromosome and pOSAK1 plasmid CDS were present. The *E. coli* Sakai pO157-cured strain showed
332 99.996 % average nucleotide identity to the Sakai chromosome (GCA_000008865.2) (95.629 %
333 alignment), with no or alignment to the pO157 plasmid, but partial coverage of pOSAK1 plasmid (100
334 % identity, 47.822 % alignment). Inoculation of *E. coli* Sakai pO157-cured with spinach plants
335 significantly reduced the number of bacteria recovered from roots and leaves compared to its
336 isogenic parent (Fig. 3B, black and white bars respectively). Binding to spinach tissue was not due to
337 generic adherence to surfaces, since there was no significant difference between the number of *E.*
338 *coli* Sakai pO157 mutant and its isogenic parent recovered from natural wool (Fig. 3B, wool grey
339 bars).

340 3.4.2 Analysis of a T2SS mutant in spinach interactions

341 To assess a role of the pO157-encoded T2SS in spinach binding, a defined knockout of the T2SS
342 secretin protein, EtpD was constructed (*E. coli* Sakai Δ *etpD*). Whole genome sequencing confirmed
343 the specific loss of the *etpD* CDS in its entirety, as designed. Average nucleotide identity between *E.*
344 *coli* Sakai Δ *etpD* and the Sakai genome (GCA_000008865.2) showed 99.997 % identity to the
345 chromosome (94.614 % alignment), and although short-read sequencing was performed, some
346 contigs covered the plasmids, with 99.960 % identity to the pO157 plasmid (49.374 % alignment).

347 Adherence of the *etpD* mutant was compared to the isogenic parent to spinach roots derived from
348 plants that were propagated in compost (Fig. 4B). Recovery of the *etpD* mutant (*E. coli* Sakai Δ *etpD*)
349 was reduced by 0.32 logCFU (95% credibility interval -0.56:-0.09) compared to the control (*E. coli*
350 Sakai WT), although adherence was not completely abrogated. Complementation of the *etpD*
351 mutant with a plasmid-borne copy of *etpD* (*E. coli* Sakai Δ *etpD* + pAH007) under inducible control did
352 not restore adherence to wild-type levels, relative to cells transformed with the empty vector
353 control (*E. coli* Sakai WT + pSE380) also treated with the inducing agent, IPTG (Fig. 4). Substantial
354 variation occurred between replicate plants and the average number of recovered bacteria with the
355 empty vector (*E. coli* Sakai WT + pSE380) was greater than the *etpD* mutant without the plasmid (*E.*
356 *coli* Sakai Δ *etpD*), indicative of an artefactual effect from the addition of IPTG. This was previously
357 reported and suggests that IPTG may influence off-target genes that directly or indirectly alter
358 adherence to plant tissue in *E. coli* Sakai [14].

359 The role in adherence for the T2SS was also tested on spinach leaf tissue to determine whether this
360 function extended to other tissue sites. Recovery of the *etpD* mutant transformed with the empty
361 vector (*E. coli* Sakai $\Delta etpD$ + pSE380) was enhanced with respect to the *etpD* mutant alone by 0.4
362 logCFU (95% credibility interval 0.15:0.63). Complementation of the *etpD* mutant using an inducible
363 version of *etpD* cloned into single-copy plasmid (*E. coli* Sakai $\Delta etpD$ + pAH007) restored binding to
364 2.6-fold greater than the *etpD* mutant (*E. coli* Sakai $\Delta etpD$ + pSE380) (Fig. 4).

365 A plant-dependent specificity for *etpD* was confirmed by assessing binding to an abiotic surface
366 (polystyrene), where there was no significant difference in attachment between *E. coli* Sakai $\Delta etpD$,
367 Sakai pO157-cured or *E. coli* Sakai WT, after either 2 hours (as measured by Crystal Violet, OD_{590nm}
368 <0.050 ±0.025 SD) or after 24 hours, in 3 different media types.

369 3.4.2 Expression of T2SS *in vitro*

370 The T2SS from *E. coli* Sakai is largely uncharacterised, both in terms of function and expression
371 profile, with no data relating to plant-relevant environments. Therefore, expression was assessed
372 from two independent plasmid-borne (multi-copy) transcriptional reporter fusions for *etpC*, the first
373 gene of the operon, and for *etpD*, the outer membrane protein, since there is 211 nt between the
374 stop codon of *etpC* and start codon of *etpD*, which includes putative transcriptional start sites (Fig.
375 5A). It appears that *etpD-K* are polycistronic since there is no apparent untranslated DNA between
376 genes, and there is a predicted ribosome binding site upstream of *etpI*. The reporter fusions
377 encompassed 508 nt and 257 nt upstream of the *etpC* and *etpD* start codons, respectively. Under *in*
378 *vitro* conditions (defined medium at 18 °C), the maximum level of expression for both genes
379 occurred in late exponential phase of growth (OD₆₀₀ ~ 1), although there were marked differences
380 in growth rates under the different carbon source regimes: *E. coli* Sakai reached this cell density in
381 two days when grown with glucose, but needed six days with glycerol as a carbon source. The
382 relative fluorescence was normalised to cell density to allow for comparison between the reporters,
383 and GFP fluorescence from the *etpD-gfp+* reporter was five- to six-fold greater than the *etpC-gfp+*
384 reporter (Fig. 5B). GFP fluorescence from both reporter constructs were three- to four-fold higher in
385 RD-MOPS glycerol compared to that in RD-MOPS glucose; indicative of catabolite repression [42].

386 3.4.4 Expression of the T2SS secretin gene, *etpD*, *in planta*

387 The transcriptional activity of the T2SS *etpD* secretin gene was assessed during *E. coli* Sakai
388 colonisation of spinach roots or leaves, using the *etpD-gfp+* transcriptional reporter plasmid
389 (pAH009). Repressive culture conditions for *etpD* expression (RD MOPS glucose: Fig. 5B white bars)
390 were used to pre-culture the cells to observe *bone fide* expression, and *E. coli* Sakai + pAH009 were
391 co-transformed with a constitutive RFP plasmid (*pmKate*) to aid location (Fig. 6). After four days, *E.*
392 *coli* Sakai + pAH009 + *pmKate* were located along the surface of intact spinach root epidermal cells

393 (Fig. 6A) or within an epidermal cell (Fig. 6B). Detection of GFP showed that *etpD* was expressed
394 both on and inside spinach root cells, and expression was heterogenous, ranging from no GFP to
395 very bright levels. Although the non-GFP expressers could have lost the reporter plasmid due to lack
396 of selective pressure, detection of RFP from *pmKate* indicated maintenance of plasmids. *E. coli* Sakai
397 located within the epidermal cell (Fig. 6Bi and ii) were apparently adherent to the plant cell wall (Fig.
398 6Biii), while others appeared to still be moving (since the plant tissue was live and unfixed during
399 imaging) (Fig. S1A, arrow). *E. coli* Sakai co-transformed with *pmKate* and a constitutive GFP reporter
400 (*pgyrA-gfp*) showed that the experimental conditions did not impact GFP detection and resulted in a
401 similar pattern of colonisation, with apparently adherent cells (Fig. S1A, circle), indicating that
402 harbouring two plasmids did not incur detrimental effects on isolate Sakai colonisation. As expected,
403 there was no GFP observed from *E. coli* Sakai co-transformed with *pmKate* and the no-promoter
404 pKC026 plasmid vector control (Fig. S1B).

405 Expression of *etpD* was also shown for endophytic *E. coli* Sakai +pAH009 + *pmKate* located within the
406 apoplast of spinach leaves (Fig. 7), from individual cells attached to spongy mesophyll cells (Fig. 7A)
407 or adjacent to the cell wall (Fig. 7B), and in small chains of cells (Fig. 7C). In contrast, no GFP was
408 observed from *E. coli* Sakai transformed with empty vector control (pKC026) (Fig. 7D).

409 4 Discussion

410 The main aim of this study was to identify novel STEC genes that mediate early interactions with
411 fresh produce plant hosts. A high-throughput positive selection approach was used, where a BAC
412 library of *E. coli* Sakai genomic fragment clones was screened for interactions to spinach roots.
413 Spinach has been linked with high profile outbreaks of STEC, and although plant roots are not
414 consumed they represent the preferred site of colonisation of *E. coli* Sakai. The screen enriched for
415 the equivalent of 2 % of the *E. coli* Sakai genome, which is in-line with other studies using alternative
416 approaches, e.g. a whole transcriptome study of *E. coli* Sakai identified two or six 'adherence' genes
417 following inoculation with lettuce plants for one hour or two days, respectively (Linden et al., 2016).
418 Several of the enriched gene loci were previously reported for STEC interaction with plant tissue,
419 validating both the screen and their potential plant-associated functional role.

420 Adherence is a key step in early interactions with host tissue and STEC fimbrial adhesins that
421 mediate specific binding to plant cell wall components include *E. coli* common pilus (ECP) and Yad
422 fimbriae [14, 43] and non-specific interactions via flagella [15]. Potential candidates enriched in the
423 screen may be involved in non-adherence functions, such as response to PAMP perception by the
424 host, Nle effectors since NleA is known to play a role in disrupting secretory pathways [44, 45] or

425 modulating host cytoskeleton (EspO1-2) [46] in animal hosts. Metabolic processes are also key for
426 colonisation, which may explain enrichment of siderophore, ChuS.

427 One of the enriched loci selected for functional assessment on the basis of potential adherence
428 included a chaperone-usher fimbrial gene cluster, termed Loc6 [11], was previously shown to be
429 induced in STEC isolate EDL933 (gene Z1536) 30 minutes after exposure to lettuce leaf lysates [19].
430 In a separate study, the gene encoding the outer membrane protein (ECs1277) was induced in *E. coli*
431 Sakai in response to a temperature reduction, to 14 °C [47]. However, the absence of any positive
432 interaction with spinach root tissue indicated either no functional role or a subtle effect on binding.
433 Alternative genes in the contiguous region identified by the BAC screen that may have contributed
434 to interactions include a two-partner secretion (TPS) system termed *otpAB* (ECs1282-1283), which
435 was characterised in STEC isolate EDL933 [48] and shares 100% sequence identity with *E. coli* Sakai.
436 Although OtpA and OtpB apparently constitute a genuine TPS system in this isolate, the gene
437 sequences did not genetically cluster with either of the two major subtypes of characterised two-
438 partner secretion systems, haemolysins or adhesins [48]. Therefore, the authors postulated that the
439 function of *otpA* could be accessory to that of the upstream fimbrial locus (*loc6*), which suggests that
440 there may be a linked function between the gene clusters.

441 The second enriched candidate region selected for functional analysis was the T2SS encoded on the
442 *E. coli* Sakai plasmid, pO157. The pO157 plasmid is ~ 93 Kb and also encodes virulence factors such
443 as haemolysin genes, a catalase, a serine protease and a toxin gene [49]. The T2SS is widespread but
444 not ubiquitous in bacteria and has been reported for bacteria from a range of hosts and
445 environmental habitats [50]. In the related phytopathogen *Pectobacterium atrosepticum*, the T2SS
446 (termed the Out system) bears structural and evolutionary similarity to the conjugative T4 pilus, and
447 the gene cluster organisation tends to be labelled with gene 'C' at the beginning and gene 'O' at the
448 end of the cluster. It is often termed the general secretory pathway (*gsp*), but in *E. coli* it is termed
449 the EHEC type II pathway (*etp*) [51]. EtpD is orthologous to the secretin protein, 'D' that forms a
450 channel across the outer member, while EtpC is homologous to the 'C' protein that spans the inner-
451 membrane as an anchoring protein [50]. Outside the *Escherichia* genus, EtpC has lower levels of
452 homology to other species T2SS than EtpD [51], but does retain the functional domain of the
453 superfamily of PulC proteins [52].

454 Absence of the pO157 plasmid reduced the number of bacteria recovered from spinach tissue, which
455 appeared to be dependent on the EtpD secretin protein. Gene expression analysis supports a role for
456 the T2SS *in planta*. The T2SS was shown to be responsive to incubation with plant tissue, with
457 induction of *etpC* in response to spinach leaf lysates and spinach root exudates, and *etpD* induced in

458 response to spinach root exudates [20]. Here, we show that expression occurs at plant-relevant
459 temperatures (18 °C), and that both *etpC* and *etpD* expression was induced in the presence of
460 glycerol but not glucose. Our data also supports independent promoter activity for both genes,
461 albeit to differing levels. It is notable that the *etp* gene cluster for *E. coli* Sakai is encoded on the
462 pO157 plasmid, whereas in other *E. coli* pathotypes the genes are chromosomal, indicative of recent
463 recombination events, which could influence regulation in a background-dependent manner. A role
464 for the STEC T2SS in colonisation of plant hosts is supported by data that shows the *etp* genes were
465 upregulated in spinach outbreak STEC isolate TW14359 compared to *E. coli* Sakai upon adherence to
466 mammalian MAC cells *in vitro* [53]. However, expression of the T2SS was not a pre-requisite for
467 colonisation of bovine GI tract [24, 54] or gnotobiotic piglet intestines [55], indicating a degree of
468 specificity in its function.

469 Whether or not the T2SS interacts directly with plant tissue, or indirectly via a T2-secreted protein, is
470 not yet clear. Functional analysis of the T2SS in STEC isolate EDL933 showed that it is required for
471 secretion of StcE (TagA), a metalloprotease that cleaves a C1-esterase inhibitor (C1-INH) [56],
472 glycoprotein 340 (gp340) and mucin7 [57]. A role for the T2SS binding to mammalian tissue was
473 demonstrated with Hep-2 cells [57], HeLa cells and in colonisation of the rabbit intestine [58].
474 Beyond that there is little available information on the STEC T2SS.

475 4.1 Conclusion

476 High-throughput screening of the *E. coli* Sakai genome, using a BAC clone library, has enabled
477 identification of a novel role for the T2SS of this foodborne pathogen. We have shown that it is
478 expressed under relevant plant-host conditions and its presence enhances the short-term
479 interactions of *E. coli* Sakai with plant hosts. Given the widespread nature of the T2SS, and a proven
480 plant-colonisation role for T2SS of phytopathogens, it is perhaps not surprising that the STEC T2SS
481 can mediate plant colonisation interactions.

482

483 **5 Acknowledgments**

484 The work was supported by a BBSRC grant to NJH, AH & LP (BB/I014179/1); Scottish Government
485 Strategic funding to NJH, PH and JM; and BBSRC Institute grant funding to DLG and SPM
486 (BB/J004227/1). Thanks to Steve Whisson for the gift of pBeloBAC11, Kathryn Wright for assistance
487 with confocal microscopy and to Jacqueline Marshall and Marta Lis for technical assistance.

488 **6 CRediT author statement**

489 **Ashleigh Holmes:** Investigation, Formal analysis, Writing - Original Draft; **Leighton Pritchard:**
490 Investigation, Formal analysis, Writing - Original Draft; **Peter Hedley:** Validation, Data Curation;
491 **Jenny Morris:** Investigation; **Sean P. McAteer:** Resources; **David L. Gally:** Resources, Funding
492 acquisition; **Nicola J. Holden:** Conceptualization, Writing - Review & Editing, Supervision, Funding
493 acquisition

494

495

496 7 References

- 497 [1] Painter JA, Hoekstra RM, Ayers T, Tauxe RV, Braden CR, Angulo FJ, et al. Attribution of
498 foodborne illnesses, hospitalizations, and deaths to food commodities by using outbreak
499 data, United States, 1998-2008. *Emerging Infectious Diseases* 2013;19(3):407-
500 15.10.3201/eid1903.111866
- 501 [2] Ceuppens S, Johannessen GS, Allende A, Tondo EC, El-Tahan F, Sampers I, et al. Risk factors
502 for *Salmonella*, Shiga toxin-producing *Escherichia coli* and *Campylobacter* occurrence in
503 primary production of leafy greens and strawberries. *International Journal of Environmental
504 Research and Public Health* 2015;12(8):9809-31.10.3390/ijerph120809809
- 505 [3] Holden N, Jackson RW, Schikora A. Plants as alternative hosts for human and animal
506 pathogens. *Frontiers in Microbiology* 2015;6:397.10.3389/fmicb.2015.00397
- 507 [4] Ibekwe AM, Grieve CM, Papiernik SK, Yang CH. Persistence of *Escherichia coli* O157:H7 on
508 the rhizosphere and phyllosphere of lettuce. *Letters in Applied Microbiology* 2009;49(6):784-
509 90
- 510 [5] Williams AP, Avery LM, Killham K, Jones DL. Survival of *Escherichia coli* O157:H7 in the
511 rhizosphere of maize grown in waste-amended soil. *Journal of Applied Microbiology*
512 2007;102(2):319-26.JAM3104 [pii] 10.1111/j.1365-2672.2006.03104.x
- 513 [6] Wright KM, Chapman S, McGeachy K, Humphris S, Campbell E, Toth IK, et al. The endophytic
514 lifestyle of *Escherichia coli* O157:H7: quantification and internal localization in roots.
515 *Phytopathology* 2013;103(4):333-40.10.1094/PHYTO-08-12-0209-FI
- 516 [7] Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM. The role of root exudates in rhizosphere
517 interactions with plants and other organisms. *Annual Review of Plant Biology* 2006;57:233-
518 66
- 519 [8] Bakker P, Berendsen RL, Doornbos RF, Wintermans PCA, Pieterse CMJ. The rhizosphere
520 revisited: root microbiomics. *Frontiers in Plant Science* 2013;4:7.10.3389/fpls.2013.00165
- 521 [9] Islam M, Doyle MP, Phatak SC, Millner P, Jiang X. Persistence of enterohemorrhagic
522 *Escherichia coli* O157:H7 in soil and on leaf lettuce and parsley grown in fields treated with
523 contaminated manure composts or irrigation water. *Journal of Food Protection*
524 2004;67(7):1365-70
- 525 [10] Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, et al. Complete genome
526 sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a
527 laboratory strain K-12. *DNA Research* 2001;8(1):11-22.10.1093/dnares/8.1.11
- 528 [11] Low AS, Holden N, Rosser T, Roe AJ, Constantinidou C, Hobman JL, et al. Analysis of fimbrial
529 gene clusters and their expression in enterohaemorrhagic *Escherichia coli* O157:H7.
530 *Environmental Microbiology* 2006;8(6):1033-47
- 531 [12] Saldana Z, Sanchez E, Xicohtencatl-Cortes J, Puente JL, Giron JA. Surface structures involved
532 in plant stomata and leaf colonization by Shiga-toxigenic *Escherichia coli* O157:H7. *Frontiers
533 in Microbiology* 2011;2:119.10.3389/fmicb.2011.00119
- 534 [13] Xicohtencatl-Cortes J, Sanchez Chacon E, Saldana Z, Freer E, Giron JA. Interaction of
535 *Escherichia coli* O157:H7 with leafy green produce. *Journal of Food Protection*
536 2009;72(7):1531-7
- 537 [14] Rossez Y, Holmes A, Lodberg-Pedersen H, Birse L, Marshall J, Willats WGT, et al. *Escherichia
538 coli* common pilus (ECP) targets arabinosyl residues in plant cell walls to mediate adhesion to
539 fresh produce plants. *Journal of Biological Chemistry* 2014;289:34349-
540 65.10.1074/jbc.M114.587717
- 541 [15] Rossez Y, Holmes A, Wolfson EB, Gally DL, Mahajan A, Pedersen HL, et al. Flagella interact
542 with ionic plant lipids to mediate adherence of pathogenic *Escherichia coli* to fresh produce
543 plants. *Environmental Microbiology* 2014;16(7):2181-95.10.1111/1462-2920.12315
- 544 [16] Shaw RK, Berger CN, Feys B, Knutton S, Pallen MJ, Frankel G. Enterohemorrhagic *Escherichia
545 coli* exploits EspA filaments for attachment to salad leaves. *Applied and Environmental
546 Microbiology* 2008;74(9):2908-14.10.1128/AEM.02704-07

- 547 [17] Wright KM, Crozier L, Marshall J, Merget B, Holmes A, Holden NJ. Differences in
548 internalization and growth of *Escherichia coli* O157:H7 within the apoplast of edible plants,
549 spinach and lettuce, compared with the model species *Nicotiana benthamiana*. *Microbial*
550 *Biotechnology* 2017;10(3):555–69.10.1111/1751-7915.12596
- 551 [18] Wendel AM, Johnson DH, Sharapov U, Grant J, Archer JR, Monson T, et al. Multistate
552 outbreak of *Escherichia coli* O157:H7 infection associated with consumption of packaged
553 spinach, August-September 2006: the Wisconsin investigation. *Clinical Infectious Diseases*
554 2009;48(8):1079-86.10.1086/597399
- 555 [19] Kyle JL, Parker CT, Goudeau D, Brandl MT. Transcriptome analysis of *Escherichia coli*
556 O157:H7 exposed to lysates of lettuce leaves. *Applied and Environmental Microbiology*
557 2010;76(5):1375-87.10.1128/AEM.02461-09
- 558 [20] Crozier L, Hedley P, Morris J, Wagstaff C, Andrews SC, Toth I, et al. Whole-transcriptome
559 analysis of verocytotoxigenic *Escherichia coli* O157:H7 (Sakai) suggests plant-species-specific
560 metabolic responses on exposure to spinach and lettuce extracts. *Frontiers in Microbiology*
561 2016;7:1088.10.3389/fmicb.2016.01088
- 562 [21] Linden IVD, Cottyn B, Uyttendaele M, Vlaemynck G, Heyndrickx M, Maes M, et al.
563 Microarray-based screening of differentially expressed genes of *E. coli* O157:H7 Sakai during
564 preharvest survival on butterhead lettuce. *Agriculture*
565 2016;6(1):6.10.3390/agriculture6010006
- 566 [22] Giddens S, Jackson R, Moon C, Jacobs M, Zhang X, Gehrig S, et al. Mutational activation of
567 niche-specific genes provides insight into regulatory networks and bacterial function in a
568 complex environment. *Proceedings of the National Academy of Sciences of the United States*
569 *of America* 2007;104:18247 - 52
- 570 [23] Bai J, McAteer SP, Paxton E, Mahajan A, Gally DL, Tree JJ. Screening of an *E. coli* O157:H7
571 bacterial artificial chromosome library by comparative genomic hybridization to identify
572 genomic regions contributing to growth in bovine gastrointestinal mucus and epithelial cell
573 colonization. *Frontiers in Microbiology* 2011;2:168.10.3389/fmicb.2011.00168
- 574 [24] Dziva F, van Diemen PM, Stevens MP, Smith AJ, Wallis TS. Identification of *Escherichia coli*
575 O157:H7 genes influencing colonization of the bovine gastrointestinal tract using signature-
576 tagged mutagenesis. *Microbiology* 2004;150(Pt 11):3631-45.10.1099/mic.0.27448-0
- 577 [25] Michino H, Araki K, Minami S, Takaya S, Sakai N, Miyazaki M, et al. Massive outbreak of
578 *Escherichia coli* O157:H7 infection in schoolchildren in Sakai City, Japan, associated with
579 consumption of white radish sprouts. *American Journal of Epidemiology* 1999;150(8):787-96
- 580 [26] Zhang Y, Laing C, Steele M, Ziebell K, Johnson R, Benson AK, et al. Genome evolution in
581 major *Escherichia coli* O157:H7 lineages. *BMC Genomics* 2007;8(1):121.10.1186/1471-2164-
582 8-121
- 583 [27] Neidhardt FC, Bloch PL, Smith DF. Culture medium for enterobacteria. *Journal of*
584 *Bacteriology* 1974;119(3):736-47.ISSN 0021-9193
- 585 [28] Tatsuno I, Horie M, Abe H, Miki T, Makino K, Shinagawa H, et al. *tox*B gene on pO157 of
586 enterohemorrhagic *Escherichia coli* O157:H7 is required for full epithelial cell adherence
587 phenotype. *Infection and Immunity* 2001;69(11):6660-9.10.1128/iai.69.11.6660-6669.2001
- 588 [29] Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30(14):2068-
589 9.10.1093/bioinformatics/btu153
- 590 [30] Cock PJ, Chilton JM, Gruning B, Johnson JE, Soranzo N. NCBI BLAST+ integrated into Galaxy.
591 *Gigascience* 2015;4:39.10.1186/s13742-015-0080-7
- 592 [31] Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, et al. The Galaxy platform for
593 accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids*
594 *Research* 2018;46(W1):W537-W44.10.1093/nar/gky379
- 595 [32] Merlin C, McAteer S, Masters M. Tools for characterization of *Escherichia coli* genes of
596 unknown function. *Journal of Bacteriology* 2002;184(16):4573-81

- 597 [33] Cherepanov PP, Wackernagel W. Gene disruption in *Escherichia coli*: TcR and KmR cassettes
598 with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene*
599 1995;158(1):9-14
- 600 [34] Merritt JH, Kadouri DE, O'Toole GA. Growing and analyzing static biofilms. *Current Protocols*
601 *in Microbiology*. 2005, p. 1B..-B..17.
- 602 [35] Shen S, Mascarenhas M, Morgan R, Rahn K, Karmali MA. Identification of four fimbria-
603 encoding genomic islands that are highly specific for verocytotoxin-producing *Escherichia*
604 *coli* serotype O157 strains. *Journal of Clinical Microbiology* 2005;43(8):3840-
605 50.10.1128/JCM.43.8.3840-3850.2005
- 606 [36] Nakano M, Iida T, Ohnishi M, Kurokawa K, Takahashi A, Tsukamoto T, et al. Association of
607 the urease gene with enterohemorrhagic *Escherichia coli* strains irrespective of their
608 serogroups. *Journal of Clinical Microbiology* 2001;39:4541 - 3
- 609 [37] Tobe T, Beatson SA, Taniguchi H, Abe H, Bailey CM, Fivian A, et al. An extensive repertoire of
610 type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their
611 dissemination. *Proceedings of the National Academy of Sciences of the United States of*
612 *America* 2006;103(40):14941-6
- 613 [38] Suits MDL, Pal GP, Nakatsu K, Matte A, Cygler M, Jia ZC. Identification of an *Escherichia coli*
614 O157:H7 heme oxygenase with tandem functional repeats. *Proceedings of the National*
615 *Academy of Sciences of the United States of America* 2005;102(47):16955-
616 60.10.1073/pnas.0504289102
- 617 [39] Landstorfer R, Simon S, Schober S, Keim D, Scherer S, Neuhaus K. Comparison of strand-
618 specific transcriptomes of enterohemorrhagic *Escherichia coli* O157:H7 EDL933 (EHEC) under
619 eleven different environmental conditions including radish sprouts and cattle feces. *BMC*
620 *Genomics* 2014;15:353.10.1186/1471-2164-15-353
- 621 [40] Lim JY, La HJ, Sheng H, Forney LJ, Hovde CJ. Influence of plasmid pO157 on *Escherichia coli*
622 O157:H7 Sakai biofilm formation. *Applied and Environmental Microbiology* 2010;76(3):963-
623 6.10.1128/aem.01068-09
- 624 [41] Jha G, Rajeshwari R, Sonti RV. Bacterial type two secretion system secreted proteins: double-
625 edged swords for plant pathogens. *Molecular Plant-Microbe Interactions* 2005;18(9):891-
626 8.10.1094/mpmi-18-0891
- 627 [42] Mayer C, Boos W. Hexose/pentose and hexitol/pentitol metabolism. *Ecosal Plus*
628 2005:doi:10.1128/ecosalplus.3.4.1.doi:10.1128/ecosalplus.3.4.1
- 629 [43] Larssonneur F, Martin FA, Mallet A, Martinez-Gil M, Semetey V, Ghigo J-M, et al. Functional
630 analysis of *Escherichia coli* Yad fimbriae reveals their potential role in environmental
631 persistence. *Environmental Microbiology* 2016;18(12):5228-48.10.1111/1462-2920.13559
- 632 [44] Gruenheid S, Sekirov I, Thomas NA, Deng WY, O'Donnell P, Goode D, et al. Identification and
633 characterization of NleA, a non-LEE-encoded type III translocated virulence factor of
634 enterohaemorrhagic *Escherichia coli* O157:H7. *Molecular Microbiology* 2004;51(5):1233-
635 49.10.1046/j.1365-2958.2004.03911.x
- 636 [45] Thanabalasuriar A, Koutsouris A, Weflen A, Mimee M, Hecht G, Gruenheid S. The bacterial
637 virulence factor NleA is required for the disruption of intestinal tight junctions by
638 enteropathogenic *Escherichia coli*. *Cellular Microbiology* 2010;12(1):31-41.10.1111/j.1462-
639 5822.2009.01376.x
- 640 [46] Morita-Ishihara T, Miura M, Iyoda S, Izumiya H, Watanabe H, Ohnishi M, et al. EspO1-2
641 regulates EspM2-mediated RhoA activity to stabilize formation of focal adhesions in
642 enterohemorrhagic *Escherichia coli*-infected host cells. *PLoS ONE*
643 2013;8(2):e55960.10.1371/journal.pone.0055960
- 644 [47] Kocharunchitt C, King T, Gobius K, Bowman JP, Ross T. Integrated transcriptomic and
645 proteomic analysis of the physiological response of *Escherichia coli* O157:H7 Sakai to steady-
646 state conditions of cold and water activity stress. *Molecular and Cellular Proteomics*
647 2012;11(1):M111.009019.10.1074/mcp.M111.009019

- 648 [48] Choi PS, Dawson AJ, Bernstein HD. Characterization of a novel two-partner secretion system
649 in *Escherichia coli* O157:H7. *Journal of Bacteriology* 2007;189(9):3452-61.10.1128/jb.01751-
650 06
- 651 [49] Makino K, Ishii K, Yasunaga T, Hattori M, Yokoyama K, Yutsudo CH, et al. Complete
652 nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic *Escherichia coli*
653 O157:H7 derived from Sakai outbreak. *DNA Research* 1998;5(1):1-9.10.1093/dnares/5.1.1
- 654 [50] Nivaskumar M, Francetic O. Type II secretion system: A magic beanstalk or a protein
655 escalator. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 2014;1843(8):1568-
656 77.<https://doi.org/10.1016/j.bbamcr.2013.12.020>
- 657 [51] Schmidt H, Henkel B, Karch H. A gene cluster closely related to type II secretion pathway
658 operons of Gram-negative bacteria is located on the large plasmid of enterohemorrhagic
659 *Escherichia coli* O157 strains. *FEMS Microbiology Letters* 1997;148(2):265-
660 72.10.1111/j.1574-6968.1997.tb10299.x
- 661 [52] Reeves PJ, Whitcombe D, Wharam S, Gibson M, Allison G, Bunce N, et al. Molecular cloning
662 and characterization of 13 out genes from *Erwinia carotovora* subspecies *carotovora*: genes
663 encoding members of a general secretion pathway (GSP) widespread in Gram-negative
664 bacteria. *Molecular Microbiology* 1993;8(3):443-56.10.1111/j.1365-2958.1993.tb01589.x
- 665 [53] Abu-Ali GS, Ouellette LM, Henderson ST, Whittam TS, Manning SD. Differences in adherence
666 and virulence gene expression between two outbreak strains of enterohaemorrhagic
667 *Escherichia coli* O157:H7. *Microbiology* 2010;156(2):408-19.10.1099/mic.0.033126-0
- 668 [54] Eckert SE, Dziva F, Chaudhuri RR, Langridge GC, Turner DJ, Pickard DJ, et al. Retrospective
669 application of transposon-directed insertion site sequencing to a library of signature-tagged
670 mini-Tn5Km2 mutants of *Escherichia coli* O157:H7 screened in cattle. *Journal of Bacteriology*
671 2011;193(7):1771-6.10.1128/jb.01292-10
- 672 [55] Tzipori S, Karch H, Wachsmuth KI, Robins-Browne RM, O'Brien AD, Lior H, et al. Role of a 60-
673 megadalton plasmid and Shiga-like toxins in the pathogenesis of infection caused by
674 enterohemorrhagic *Escherichia coli* O157:H7 in gnotobiotic piglets. *Infection and Immunity*
675 1987;55(12):3117-25
- 676 [56] Lathem WW, Grys TE, Witowski SE, Torres AG, Kaper JB, Tarr PI, et al. StcE, a
677 metalloprotease secreted by *Escherichia coli* O157:H7, specifically cleaves C1 esterase
678 inhibitor. *Molecular Microbiology* 2002;45(2):277-88
- 679 [57] Grys TE, Siegel MB, Lathem WW, Welch RA. The StcE protease contributes to intimate
680 adherence of enterohemorrhagic *Escherichia coli* O157:H7 to host cells. *Infection and*
681 *Immunity* 2005;73(3):1295-303.10.1128/iai.73.3.1295-1303.2005
- 682 [58] Ho TD, Davis BM, Ritchie JM, Waldor MK. Type 2 secretion promotes enterohemorrhagic
683 *Escherichia coli* adherence and intestinal colonization. *Infection and Immunity*
684 2008;76(5):1858-65.10.1128/IAI.01688-07

685

686

687 **8 Tables**

688 **Table S1** Description of gene loci enriched by the adherence screen, indicating S-loop and
689 genomic location; output data from the enrichment analysis; gene annotation and description.

690 **Table S2** Description of plasmid and primers used in the study.

691

692 **9 Figure Legends**

693 **Figure 1** Regions in *E. coli* Sakai genome enriched by the adherence screen

694 Output from the model indicating estimated values of delta: the effect of treatment (passage) on
695 retention of the introduced *E. coli* Sakai DNA in an *E. coli* DH10B background, for (A) Sakai
696 chromosome DNA; (B) Sakai pO157 DNA; (C) the DH10B chromosome background. Estimated values
697 are shown as black dots, and the 50 % credibility interval (CI) of this value as a vertical line. Where
698 the 50 % CI does not include the median value for the dataset (assumed to represent a neutral
699 response to passage), this may imply a selection response. Green CIs, where the median response is
700 lower than the 50 % CI, are interpreted as positive selection pressure such that the gene is beneficial
701 under passage. Magenta CIs, where the median response is greater than the 50 % CI, are interpreted
702 as negative selection pressure such that the gene is deleterious under passage. Regions of the *E. coli*
703 Sakai genome that are potentially under positive selection pressure include S-loop 71, S-loop 231,
704 and S-loop 225; SpLE1, and the plasmid genes encoding the Etp type II secretion system, and StcE, as
705 indicated. The DH10B chromosome genes show no evidence of positive or negative selection. Gene
706 loci are listed in Table S1.

707 **Figure 2** Assessment of *E. coli* Sakai Loc6 fimbriae in binding to spinach root tissue

708 *E. coli* Sakai or its isogenic *loc6* mutant recovered after a 2 hour adherence assay on spinach roots.
709 The data from 3 independent experiments with 10 biological replicates for each bacterial strain are
710 presented in box plots with the mean shown as a line in the interquartile ranges, and whiskers for
711 maximum and minimum values. There was no statistically significant difference in the mean number
712 of *E. coli* Sakai WT recovered compared to $\Delta loc6$ by Students *t* test ($p=0.3268$)

713 **Figure 3** *E. coli* Sakai pO157 mediates interactions with spinach tissues

714 (A) *E. coli* DH₁₀B transformed with BAC clone BAC2B5, containing pO157 sequence, or the empty BAC
715 vector pV41 recovered from roots of hydroponics-grown spinach (filled bars) or natural wool (striped
716 bars) and (B) *E. coli* Sakai WT or pO157-cured recovered from roots of compost-grown spinach (filled
717 bars), leaves (open bars) or natural wool (striped bars). Data shown is the average from triplicate

718 experiments each with five biological replicates. Statistical significance was calculated by students *t*
719 test (* $p < 0.05$, NS not significant).

720 **Figure 4** Modelling the impact of *E. coli* Sakai T2SS in interactions with spinach leaves and
721 roots

722 Bacteria recovered from spinach plant tissue after 2 hour adherence assay. Regression coefficients
723 (parameter estimates) obtained when fitting recovery data (CFU) from *E. coli* Sakai WT, $\Delta etpD$ and
724 *etpD* mutant complemented with pSE380 or pAH007 (pSE_*etpD*) under IPTG-induction from leaves
725 (A) or roots (B) to a linear model of additive effects, for each tissue. Sakai: expected recovery
726 (logCFU) of wild-type *E. coli* Sakai; Sakai $\Delta etpD$: expected (differential) effect on recovery of $\Delta etpD$
727 knockout with respect to wild-type Sakai; $\Delta etpD$ pSE380: expected (differential) effect on recovery of
728 introducing the pSE380 into the knockout background; $\Delta etpD$ pSE_EtpD: expected (differential)
729 effect on recovery of expressing EtpD, with respect to pSE380 alone. For each estimate, the marker
730 represents the median value, and vertical lines represent the extent of the 50% credibility interval
731 (50% of runs produce a value within this range).

732 **Figure 5** The *E. coli* Sakai *etp* T2SS operon and *in vitro* expression at 18 °C

733 Genetic organisation of the *etp* operon including the upstream metalloprotease gene *stcE* (A). GFP
734 reporter activity for gene expression from the 5'UTR of *etpC* (508 bp) or *etpD* (211 bp) in *E. coli*
735 Sakai, grown in RD MOPS medium supplemented with glucose (white) or glycerol (black). Expression
736 values were corrected for background from the promoter-less reporter plasmid (pKC026) measured
737 at the same optical density, and RFU normalised for cell density (OD₆₀₀). Equivalent expression levels
738 at late-exponential phase are provided (OD_{600nm} of 1) from two experimental repeats.

739 **Figure 6** Expression of *E. coli* Sakai *etpD* during root colonisation

740 Spinach roots inoculated with 10⁸ cfu of *E. coli* Sakai co-transformed with pmKate and pAH009
741 (*etpD-gfp+*) were imaged by confocal microscopy after 4 days. *E. coli* Sakai were located along (A) or
742 within (B) root epidermal cells with some *E. coli* Sakai attached to the cell wall within an epidermal
743 cell (Biii). Maximum intensity projections (A, Bi-ii) of root epidermal cells with the merged image (Ai,
744 Bi) or green channel (Aii, Bii-iii). GFP expression in green and RFP expression in magenta; root cell
745 wall autofluorescence is also detected in the magenta channel (Ai). Scale bars are 10 μm . The panel
746 of images are representative of four independent experiments from individual plants.

747 **Figure 7** Expression of *E. coli* Sakai *etpD* in spinach leaves

748 Spinach leaves infiltrated with *E. coli* Sakai co-transformed with pmKate (constitutive expression of
749 RFP) and pAH009 (*etpD-gfp+*) (A-C) or promoterless *gfp+* vector pKC026 (D) were imaged by confocal

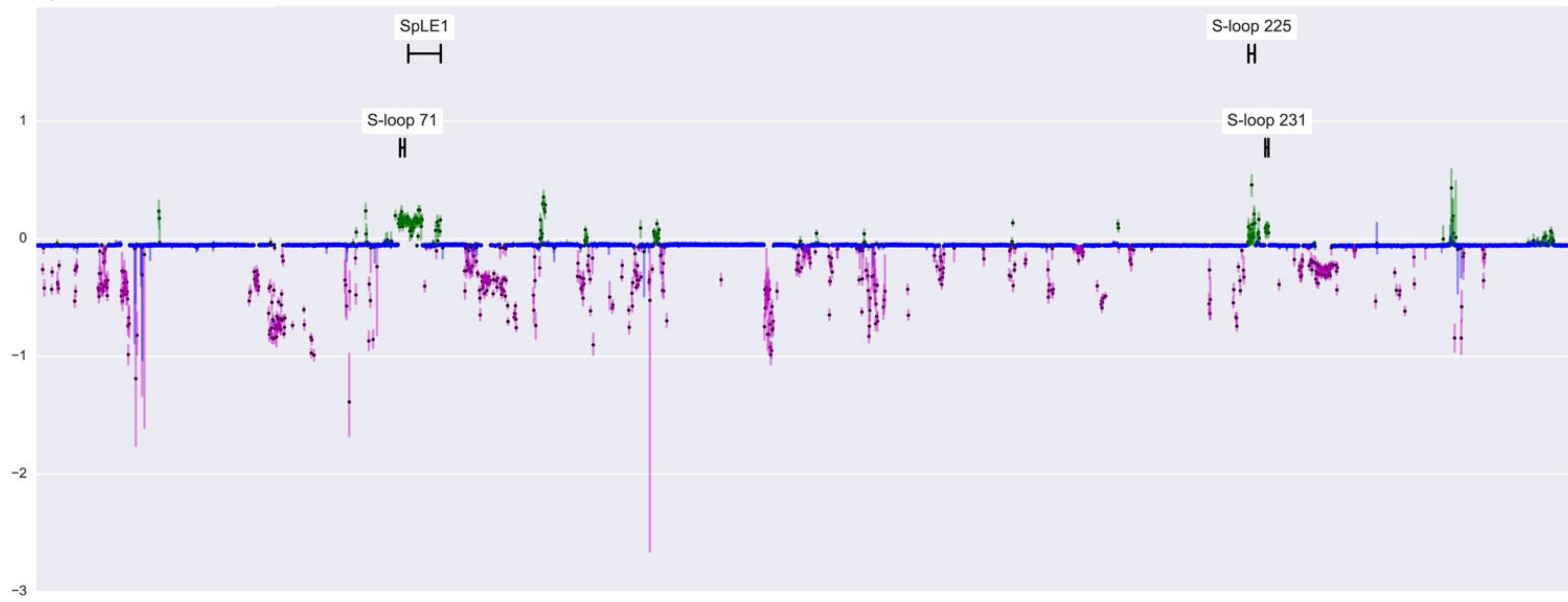
750 microscopy after four days. Chloroplast autofluorescence is false coloured blue in the images; GFP
751 expression in green and RFP expression in magenta. Three sets of parallel panels show the maximum
752 intensity projection of abaxial epidermal and mesophyll cells with the merged image (left), green
753 channel (centre) and red channel (right). The panel of images are representative of two independent
754 experiments from individual plants. Scale bars are 25 μm (A) or 5 μm (B-D). Examples of co-
755 expression of *etpD-gfp* and *rfp* are indicated by white arrows (B).

756 **Supplementary Figure 1** Confocal microscopy controls of spinach root colonisation

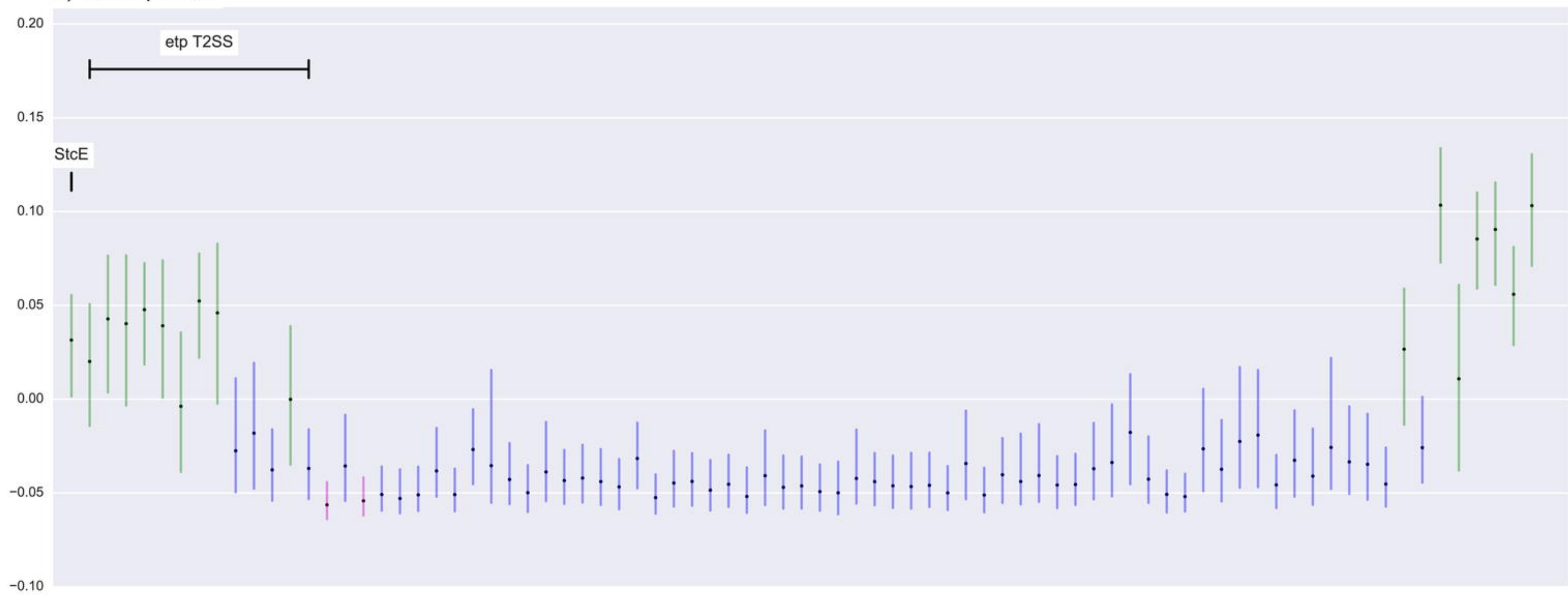
757 Spinach roots inoculated with 10^8 cfu of *E. coli* Sakai co-transformed with pmKate and *pgyrA-gfp+* (A)
758 or empty vector pKC026 (B) were imaged by confocal microscopy after 4 days. Maximum intensity
759 projections (left column A and B) of spinach root epidermal cells colonised by *E. coli* Sakai, within (A)
760 or on the surface (B) of the cell. Volume projection of spinach root epidermal cell (A right column)
761 showing *E. coli* Sakai colonisation within the epidermal cell with bacteria attached to the plant cell
762 wall (circled). *E. coli* Sakai may also have been moving during image acquisition (arrow). GFP
763 expression is coloured green and RFP expression in magenta. Scale bars are 10 μm . The panel of
764 images are representative of four independent experiments from individual plants.

765

a) Sakai chromosome

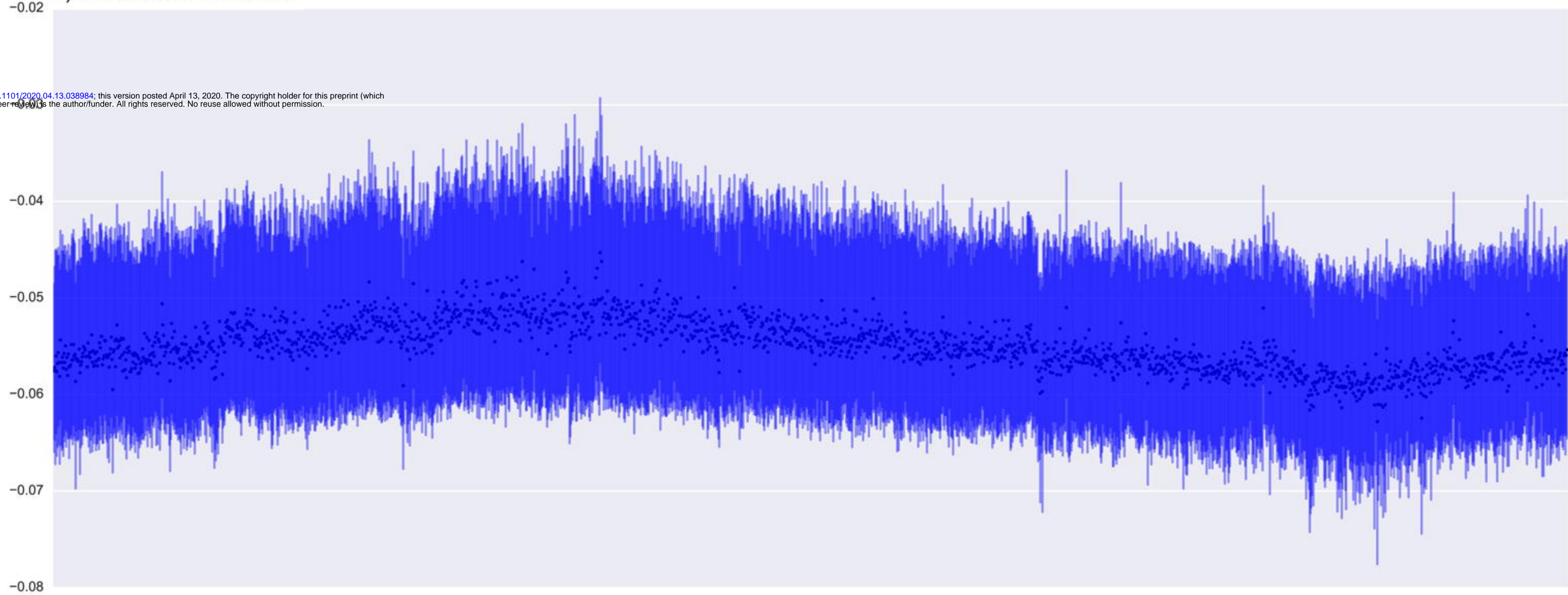


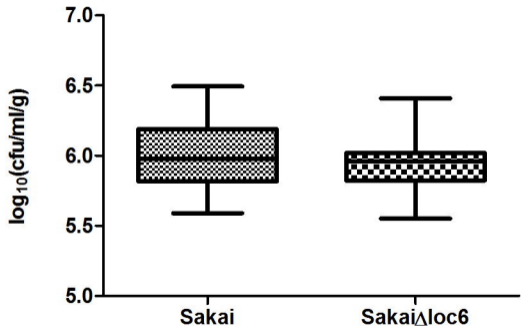
b) Sakai pO157



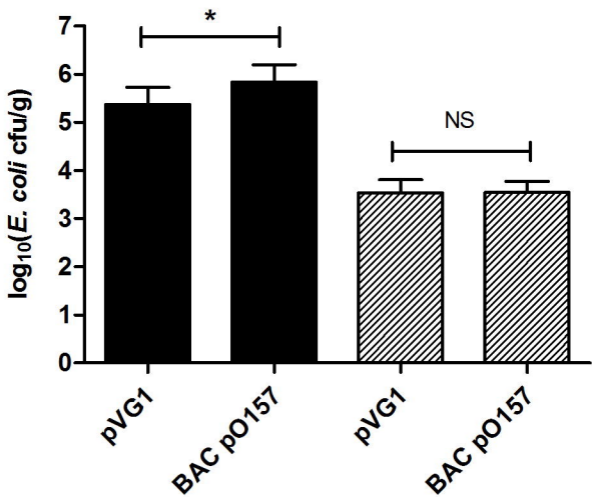
c) DH10B chromosome

bioRxiv preprint doi: <https://doi.org/10.1101/2020.04.13.038984>; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

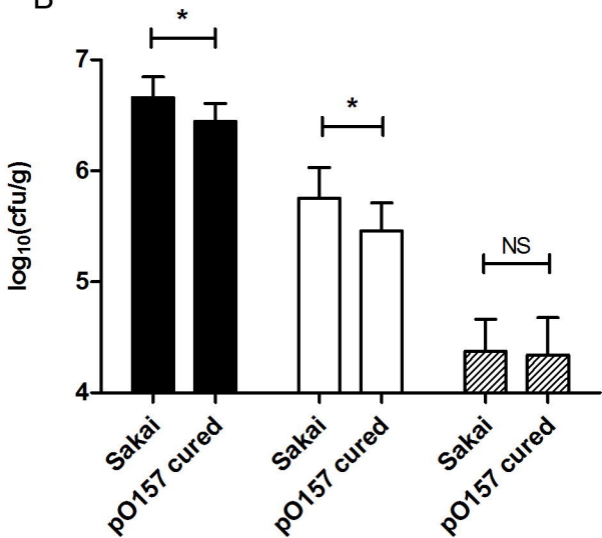


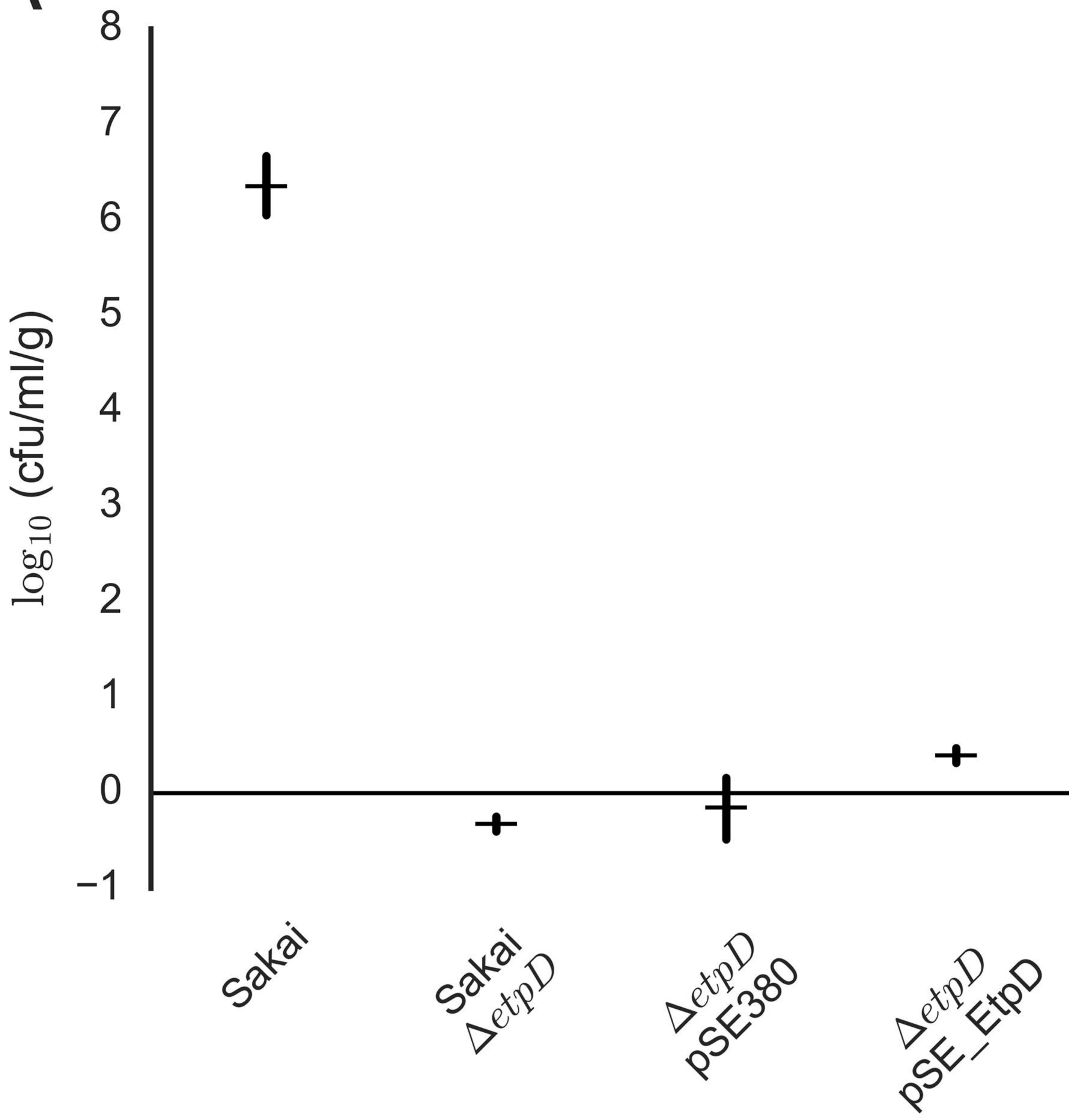
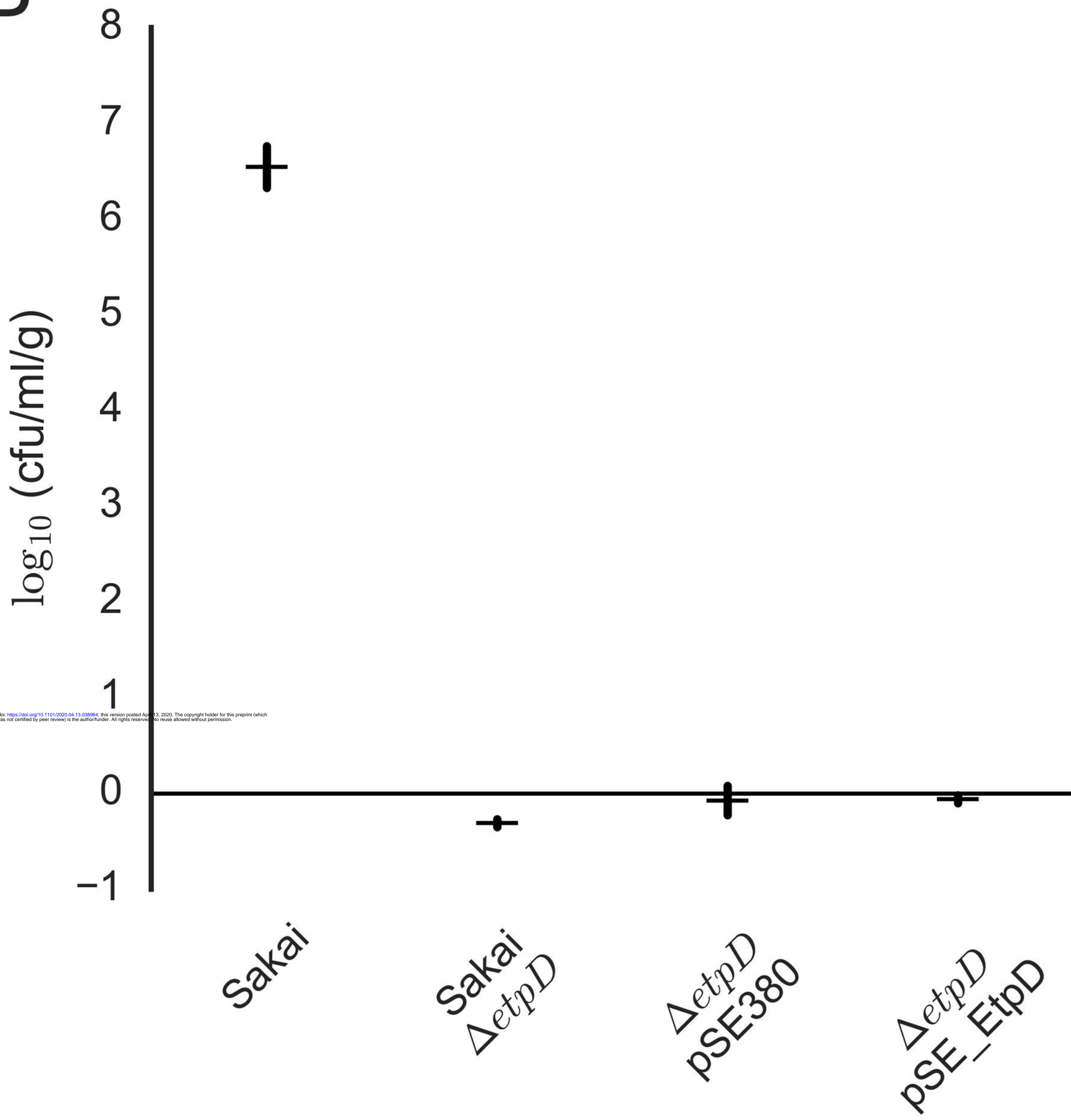


A



B



A**B**

A**B**