

1 Adding function to the genome of African *Salmonella* ST313

2 Rocío Canals¹, Disa L. Hammarlöf¹, Carsten Kröger¹, Siân V. Owen¹, Wai Yee Fong¹, Lizeth
3 Lacharme-Lora¹, Xiaojun Zhu¹, Nicolas Wenner¹, Sarah E. Carden², Jared Honeycutt², Denise
4 M. Monack², Robert A. Kingsley³, Philip Brownridge¹, Roy R. Chaudhuri⁴, Will P. M. Rowe¹,
5 Alexander V. Predeus¹, Karsten Hokamp⁵, Melita A. Gordon^{6,7}, and Jay C. D. Hinton^{1,*}

6 ¹ Institute of Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, United Kingdom

7 ² Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA
8 94305, USA

9 ³ Quadram Institute Bioscience, Norwich Research Park, Norwich, NR4 7UA, United Kingdom

10 ⁴ Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, S10 2TN,
11 United Kingdom

12 ⁵ Department of Genetics, School of Genetics and Microbiology, Smurfit Institute of Genetics, Trinity
13 College Dublin, Dublin 2, Ireland

14 ⁶ Institute of Infection and Global Health, University of Liverpool, Liverpool, L69 7ZB, United Kingdom

15 ⁷ Malawi-Liverpool-Wellcome Trust Clinical Research Programme, University of Malawi College of
16 Medicine, Blantyre 3, Malawi, Central Africa

17 * Correspondence: jay.hinton@liverpool.ac.uk

18 **Current address:**

19 Disa L. Hammarlöf, Department of Cell and Molecular Biology, Uppsala University, Uppsala, 751 24,
20 Sweden

21 Carsten Kröger, Department of Microbiology, School of Genetics and Microbiology, Moyne Institute of
22 Preventive Medicine, Trinity College Dublin, Dublin 2, Ireland

23 Siân V. Owen, Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts,
24 02115, USA

25 Will P. M. Rowe, Scientific Computing Department, STFC Daresbury Laboratory, Warrington, WA4
26 4AD, United Kingdom

27 **Abstract**

28 *Salmonella* Typhimurium ST313 causes invasive nontyphoidal *Salmonella* (iNTS) disease in sub-
29 Saharan Africa, targeting susceptible HIV⁺, malarial or malnourished individuals. An in-depth genomic
30 comparison between the ST313 isolate D23580, and the well-characterized ST19 isolate 4/74 that
31 causes gastroenteritis across the globe, revealed extensive synteny. To understand how the 856
32 nucleotide variations generated phenotypic differences, we devised a large-scale experimental
33 approach that involved the global gene expression analysis of strains D23580 and 4/74 grown in sixteen
34 infection-relevant growth conditions. Comparison of transcriptional patterns identified virulence and
35 metabolic genes that were differentially expressed between D23580 versus 4/74, many of which were
36 validated by proteomics. We also uncovered the *S. Typhimurium* D23580 and 4/74 genes that showed
37 expression differences during infection of murine macrophages. Our comparative transcriptomic data
38 are presented in a new enhanced version of the *Salmonella* expression compendium SalComD23580:
39 bioinf.gen.tcd.ie/cgi-bin/salcom_v2.pl. We discovered that the ablation of melibiose utilization was
40 caused by 3 independent SNP mutations in D23580 that are shared across ST313 lineage 2, suggesting
41 that the ability to catabolise this carbon source has been negatively selected during ST313 evolution.
42 The data revealed a novel plasmid maintenance system involving a plasmid-encoded CysS cysteinyl-
43 tRNA synthetase, highlighting the power of large-scale comparative multi-condition analyses to pinpoint
44 key phenotypic differences between bacterial pathovariants.

45 **Introduction**

46 *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*) infects a wide range of animal hosts, and
47 generally causes self-limiting gastroenteritis in humans. Variants of this serovar, belonging to
48 sequence-type ST313, are associated with invasive nontyphoidal *Salmonella* (iNTS) disease in
49 susceptible HIV⁺, malaria-infected or malnourished individuals in sub-Saharan Africa [1]. iNTS causes
50 around 681,000 deaths per year worldwide, killing 388,000 people in Africa alone [2]. The multidrug
51 resistance of ST313 isolates complicates patient treatment and accounts for the high case fatality rate
52 (20.6%) of iNTS disease [3]. Two ST313 lineages have been associated with iNTS, and the clonal
53 replacement of lineage 1 by lineage 2 is hypothesized to have been driven by the gain of
54 chloramphenicol resistance by lineage 2 [4]. Genetically-distinct ST313 isolates that do not belong to
55 lineages 1 and 2 have been described in the United Kingdom [5] and in Brazil [6].

56 The globally-distributed *S. Typhimurium* sequence type ST19 causes gastroenteritis in humans and
57 invasive disease in mice. Following oral ingestion, these bacteria colonise the gut and stimulate
58 inflammation by a *Salmonella* pathogenicity island (SPI)-1-mediated process. Subsequently, ST19 can
59 survive, and proliferate in a “*Salmonella*-containing vacuole” (SCV) within epithelial cells or
60 macrophages that involves the SPI-2 type three secretion system responsible for systemic disease in
61 mammalian hosts [7]. Host-restriction of other *Salmonella* pathovariants has been associated with
62 genome degradation caused by pseudogene formation [8–11]. This process involves the loss or
63 inactivation of virulence genes required for colonisation of the mammalian gut, whilst the ability to thrive
64 inside macrophages is maintained.

65 Phenotypic differences between ST313 and ST19 have been summarized previously [12], and new
66 studies have since been published. [S1 Table](#) lists 20 phenotypic features that differentiate ST313 from
67 ST19 isolates, at the level of metabolism, motility, and stress resistance [13–24]. In terms of infection
68 biology, reports of the relative ability of ST313 and ST19 isolates to invade epithelial cells and
69 macrophages have yielded conflicting results ([S1 Table](#)) [6,13,15,17,25–27]. It is clear that ST313
70 infection of macrophages stimulates lower levels of cytotoxicity and inflammasome response than ST19
71 infections [13,25]. Following treatment with human serum, more complement was required for antibody-
72 mediated bactericidal killing of ST19 than for ST313 isolates [14]. Animal infection experiments have
73 demonstrated that ST313 isolates can infect non-human hosts, including mice, cows, chickens and
74 macaques [15–18,28,29]. Taken together, these findings confirm that ST313 is a distinct pathovariant
75 of *S. Typhimurium* [30]. However, the molecular mechanisms responsible for the phenotypic signature
76 of ST313 pathovariant remain to be understood, and required a bespoke experimental approach.

77 D23580 is the ST313 lineage 2 reference strain, a typical representative Malawian isolate isolated
78 from an HIV-negative child in 2004 [19]. We previously defined the transcription start sites (TSS) of this
79 strain, and identified a SNP in the promoter of the *pgtE* gene specific to ST313 lineage 2 that modulated
80 virulence [20]. To investigate whether the ability of the ST313 and ST19 sequence types of *S.*
81 *Typhimurium* to cause different types of human disease was a genetic characteristic of the two types
82 of bacteria, we identified all genomic differences between D23580 and 4/74. We then generated a
83 comprehensive dataset for studying the mechanisms of infection-relevant differences between ST313
84 and ST19 listed in [S1 Table](#). We hypothesised that transcriptional differences between the two strains

85 would account for specific phenotypic differences and we present a multi-condition transcriptomic
86 comparison of the ST313 strain, D23580, with the ST19 strain, 4/74 (S1 Fig).

87 **Results**

88 **Resequencing and reannotation of D23580, the *Salmonella* Typhimurium ST313 reference strain**

89 *S. Typhimurium* D23580 was the first ST313 isolate to be genome sequenced [19]. At that time, the
90 presence of one D23580-specific plasmid, pBT1, was reported. To facilitate a robust transcriptomic
91 analysis of D23580, we re-sequenced the strain using a combination of long-read PacBio and short-
92 read Illumina technologies. Following a hybrid assembly approach (Materials and Methods), three
93 contigs were identified: the 4,879,402 bp chromosome, the 117,046 bp pSLT-BT plasmid, and the
94 84,543 bp pBT1 plasmid (accession: XXXXXXXXXX). Comparison with the published D23580 genome
95 (accession: FN424405) [19], identified just three nucleotide differences in the chromosome.
96 Specifically, an extra nucleotide at the 304,327 position (1 bp downstream of Asp-tRNA), at the 857,583
97 position (1 bp upstream of Lys-tRNA), and one nucleotide change at position 75,492 (T-to-C; intergenic
98 region) were identified. The sequence of the pSLT-BT plasmid had a single-nucleotide deletion
99 difference at position 473, in an intergenic region. The sequence of the pBT1 plasmid has not been
100 reported previously, and a primer-walking approach was used to sequence the two remaining small
101 plasmids carried by D23580 (Materials and Methods), pBT2 and pBT3 (2556 bp and 1975 bp,
102 respectively) (accession: XXXXXXXXXX).

103 To maximise the functional insights to be gained from a transcriptomic analysis, a well-annotated
104 genome is required. The published annotation for D23580 dates back to 2009 [19], and lacked certain
105 essential bacterial genes such as the two outer membrane proteins *lppA* and *lppB* [31]. Accordingly,
106 we searched for important non-annotated bacterial genes, and used D23580 transcriptomic data
107 (described below) to cross-reference the locations of transcripts with the location of coding genes (S1
108 Text). This analysis allowed us to update the published annotation of D23580 by adding 86 new coding
109 genes, 287 sRNAs, and correcting the start or end locations of 13 coding genes (S2 Table). The re-
110 sequenced and re-annotated *S. Typhimurium* D23580 genome is subsequently referred to as
111 D23580_liv (accession: XXXXXXXXXX).

112 **The *S. Typhimurium* D23580 and 4/74 genomes are 95% identical**

113 Previously the D23580 genome had been compared with the attenuated laboratory *S. Typhimurium*
114 LT2 strain [19,32,33]. To assess the similarities and differences between the ST313 strain D23580 and
115 a virulent ST19 isolate, a detailed comparative genomic analysis was performed against the ST19 strain
116 4/74 (S1 Text). 4/74 is a prototrophic *S. Typhimurium* ST19 strain that is highly virulent in four animal
117 models [34], and is the parent of the widely-used SL1344 auxotrophic strain [35]. D23580 and 4/74
118 share 92% and 95% of coding genes and sRNAs, respectively (S2 Table). Genetic differences included
119 788 SNPs, 3 multi-nucleotide polymorphisms (MNPs), 65 indels, as well as 77 D23580-specific
120 pseudogenes that have been listed elsewhere [19]. Analysis of the SNPs, using the 4/74 annotation as
121 a reference, showed that 379 were non-synonymous, 255 were synonymous, six were located in
122 sRNAs, nine generated stop codons in coding genes and seven inactivated stop codons in D23580 (S3
123 Table). The final 132 SNPs were in intergenic regions. Fig 1 compares the chromosome and pSLT
124 plasmid organisation of strains 4/74 and D23580, and shows the distribution of the indels and three
125 SNP classes that differentiate the two strains. Seventeen of the SNPs and indels were located \leq 40
126 nucleotides upstream of one of the D23580 TSS [20], raising the possibility of a direct influence upon
127 the level of transcription.

128 Regarding prophage complement, SopE ϕ [36] was absent from D23580 and present in strain 4/74
129 [19,24]. As we established earlier, D23580 carries two ST313-specific prophages, BTP1 and BTP5
130 [19,24]. In terms of plasmids, the genome of 4/74 includes pSLT^{4/74}, pCol1B9^{4/74}, and pRSF1010^{4/74}
131 [35]. In contrast, D23580 carries a distinct plasmid complement, namely pSLT-BT, pBT1, pBT2 and
132 pBT3 [19]. The pSLT-BT plasmid of D23580 carries a Tn21-based insertion element that encodes
133 resistance to five antibiotics [19].

134 The D23580 and 4/74 strains carry 4396 orthologous coding genes (S1 Text). Ten of the orthologues
135 were encoded by the D23580-specific prophages BTP1 and BTP5, or by the 4/74-specific pRSF1010^{4/74}
136 plasmid, and so were excluded from further analysis. A total of 279 orthologous sRNAs were found in
137 both strains. The sRNA-associated differences included three 4/74-specific sRNAs (STnc3640,
138 STnc1400, STnc3800), and the duplication of *IsrB-1* in D23580. Eight new sRNAs were found in the
139 BTP1 prophage region of D23580, and the existence of four was confirmed by northern blot (S2 Fig).

140 We identified 93 D23580-specific chromosomal genes that were encoded within prophage regions
141 and absent from 4/74 (S2 Table): specifically, 59 BTP1 genes, 27 BTP5 genes, one Gifsy-2 gene and
142 six Gifsy-1 genes. We found 89 4/74-specific chromosomal genes that were absent from D23580 (S2

143 [Table](#)). Most were located in the SopE ϕ prophage region (68 genes), the Def2 remnant phage (13
144 genes), or three separate non-phage associated regions in D23580: *allB* (associated with allantoin
145 utilization); the SPI-5 genes *orfX* and *SL1344_1032*; and an approximately 4 kb deletion that included
146 genes *SL1344_1478* to *SL1344_82*.

147 A total of 4675 orthologous coding genes and non-coding sRNAs were shared by strains D23580
148 and 4/74. The sRNA *IsrB-1* was removed as it was duplicated in D23580. To search for a distinct
149 transcriptional signature of D23580, the expression levels of the 4674 orthologs was compared between
150 D23580 and 4/74 using a transcriptomic approach.

151 **Comparison of transcriptional response to infection-relevant stress between *S. Typhimurium*** 152 **ST313 D23580 and ST19 strain 4/74**

153 To discover the similarities and differences in the transcriptome of strains D23580 and 4/74, we first
154 used our established experimental strategy: the transcriptome of D23580 was determined using RNA
155 isolated from 16 infection-relevant *in vitro* growth conditions [37], and during intra-macrophage infection
156 [38,39]. To allow direct comparison of the D23580 transcriptomic data with strain 4/74, experiments
157 were performed exactly as Kröger *et al.* (2013) and Srikumar *et al.* (2015) (Materials and Methods)
158 [37,39].

159 The RNA-seq-derived sequence reads were mapped to the D23580_liv chromosome, and the pSLT-
160 BT, pBT1, pBT2 and pBT3 plasmid sequences (Materials and Methods). Numbers of mapped sequence
161 reads and other RNA-seq-derived statistical information are detailed in [S4 Table](#). The level of
162 expression of individual genes and sRNAs was calculated as transcripts per million (TPM) [40,41] for
163 the chromosome, and the pSLT-BT and pBT1 plasmids ([S5 Table](#)). To achieve a complete
164 transcriptomic comparison, we first re-analysed our 4/74 transcriptomic data [37,39] to add all
165 transcripts expressed by the three plasmids, pSLT^{4/74}, pCol1B9^{4/74}, and pRSF1010^{4/74} (Materials and
166 Methods, [S4 Table](#)).

167 Initial analysis focused on the expression characteristics of the strains D23580 and 4/74 in 17 distinct
168 environmental conditions. The number of genes and sRNAs expressed in at least one condition for
169 strain D23580 was 4365 (85%) out of 5110. 745 genes and sRNAs (15%) were not expressed in any
170 of the 17 conditions. For strain 4/74, the number of genes and sRNAs that were expressed in at least
171 one condition was 4306 (86%) out of 5026, consistent with our earlier findings [37] ([S3 Fig](#)). 3958 of the

172 4674 orthologous coding genes and sRNAs shared by strains D23580 and 4/74 were expressed in at
173 least one growth condition in both strains.

174 A small minority (117) of orthologous genes were expressed in at least one condition in strain 4/74,
175 but not in any of the conditions in D23580, with most showing low levels of expression (close to the
176 threshold TPM = 10) ([S5 Table](#)). In contrast, we identified 82 orthologous coding genes and sRNAs that
177 were expressed in at least one of the 17 growth conditions for D23580, but not expressed in 4/74 ([S5](#)
178 [Table](#)).

179 To compare the expression profiles of D23580 and 4/74, we made 17 individual pair-wise
180 comparisons between the 17 growth conditions with the two strains (Materials and Methods, [S3 Fig](#)).
181 The data confirmed that *S. Typhimurium* reacts to particular infection-relevant stresses with a series of
182 defined transcriptional programmes that we detailed previously [37]. By comparing the transcriptomic
183 response of two pathovariants of *S. Typhimurium*, the conservation of the transcriptional response is
184 apparent ([S3 Fig](#)).

185 A complementary analytical approach was used to identify the transcriptional differences that relate
186 to the distinct phenotypes of the ST313 and ST19 pathovariants ([S1 Table](#)). Overall, 1031 of the
187 orthologous coding genes and sRNAs were differentially expressed (≥ 3 fold-change) between strains
188 D23580 and 4/74, in at least one growth condition ([Fig 2A](#)). Transcriptional differences are highlighted
189 in [S3 Fig](#).

190 The terms “D23580-upregulated” and “D23580-downregulated” refer to genes that are either more
191 or less expressed in D23580, compared to 4/74. Three coding genes were D23580-upregulated and
192 six genes were D23580-downregulated in almost all growth conditions ([Fig 2B and 2C](#)). The up-
193 regulated genes included *pgtE*, a gene that is highly expressed in D23580, responsible for resistance
194 to human serum killing and linked to virulence [20]. The other two upregulated genes were *nlp*, encoding
195 a *ner*-like regulatory protein, and the *STM2475* (*SL1344_2438*) gene which encodes a hypothetical
196 protein.

197 Three of the genes that were D23580-downregulated in most conditions (*pSLT043-5*) were located
198 downstream of the Tn21-like element in the pSLT-BT plasmid ([S4 Fig](#)). Because the Tn21-like multidrug
199 resistance island was inserted between the *mig-5* promoter region and the *pSLT043-5* genes, we
200 hypothesise that the differential expression reflects transcriptional termination mediated by the Tn21

201 cassette. Two other D23580-downregulated genes were located in the Gifsy-1 prophage region, *dinI-*
202 *gfoA*. The presence of a SNP in the promoter D23580 $P_{dinI-gfoA}$ has already been proven to be
203 responsible for the lack of viability of the Gifsy-1 phage in D23580 [24]. The final gene that was D23580-
204 downregulated in most growth conditions was the *cysS* chromosomal gene, which encodes a cysteinyl-
205 tRNA synthetase. Aminoacyl-tRNA synthetases are generally essential genes, required for cell growth
206 and survival [42]. The unexpected low level of *cysS* expression in D23580 in several growth conditions
207 (TPM values ranging from 5 to 18 excluding the late stationary phase and shock conditions) was
208 investigated further (see below).

209 Intriguing patterns of differential expression were observed between strains D23580 and 4/74 in
210 particular growth conditions for certain functional groups of *Salmonella* genes. For example, the flagellar
211 regulon and associated genes showed a characteristic pattern of expression in the phosphate carbon
212 nitrogen (PCN)-related minimal media and inside macrophages (S5 Fig). To allow us to make
213 statistically significant findings, a larger-scale experiment was designed.

214 **Identification of the transcriptional signature of *S. Typhimurium* ST313 D23580**

215 To generate a robust transcriptional signature of D23580, we focused on the five environmental
216 conditions with particular relevance to *Salmonella* virulence, namely ESP (early stationary phase),
217 anaerobic growth, NonSPI2 (SPI2-non-inducing), InSPI2 (SPI2-inducing) conditions and intra-
218 macrophage. The ESP and anaerobic growth conditions stimulate expression of the SPI-1 virulence
219 system, and SPI-2 expression is induced by the InSPI2 and macrophage conditions [37,39]. RNA was
220 isolated from three biological replicates of both D23580 and 4/74 grown in the four *in vitro* environmental
221 conditions. The three biological replicates were generated in parallel, in a new set of experiments.
222 Additionally, RNA was extracted from two additional biological replicates of intra-macrophage *S.*
223 *Typhimurium*, following infection of murine RAW264.7 macrophages for both D23580 and 4/74.
224 Following RNA-seq, the sequence reads were mapped to the D23580 and 4/74 genomes using our
225 bespoke software pipeline (Materials and Methods). The RNA-seq mapping statistics are detailed in S4
226 Table. To ensure that biologically meaningful gene expression differences were reported, we used very
227 conservative cut-offs to define differential expression (Materials and Methods).

228 Following RNA-seq analysis of the three biological replicates of D23580 and 4/74 in five growth
229 conditions, differential expression analysis of orthologous genes and sRNAs was performed with a

230 rigorous statistical approach (Materials and Methods, [S6 Table](#)). We identified 677 genes and sRNAs
231 that showed ≥ 2 fold-change ($FDR \leq 0.001$) in at least one growth condition ([Fig 3A](#)). Between 6%
232 (anaerobic growth) and 2% (InSPI2 condition) of orthologous genes and sRNAs were differentially
233 expressed between the two strains ([Fig 3A and 3B](#)).

234 The ability to swim in semi-solid agar is a key phenotypic difference between D23580 and 4/74
235 [13,16]. We confirmed that D23580 was less motile than 4/74 ([S5 Fig](#)), but did not observe significant
236 differences in motility gene expression in complex media between the strains at the transcriptional level
237 ([S5 Fig](#)). One nucleotide deletion and 11 SNP differences were found in the flagellar regulon between
238 the two strains: one mutation in the promoter region of *mcpA*; three synonymous mutations in *flgK*,
239 *cheA*, and *fliP*; four non-synonymous mutations in *flhA*, *flhB*, *fliB*, and *mcpC*; and three mutations in the
240 5'UTRs of *motA*, *flhD*, and *mcpA*. FlhA and FlhB are transmembrane proteins that are essential for
241 flagellar protein export [43]. The SNP in *flhB* is 4/74-specific, as other ST19 strains, such as LT2 and
242 14028, conserved the SNP in D23580. The D23580 *flhA* SNP was specific to ST313 lineage 2.

243 To investigate the function of the 4/74 *flhA* SNP, the mutation was introduced to the chromosome of
244 D23580 by single nucleotide engineering (D23580 *flhA*^{4/74}). Motility of the D23580 *flhA*^{4/74} mutant was
245 significantly increased compared to the D23580 wild-type strain ([S5 Fig](#)). We originally hypothesized
246 that the *flhA* SNP was related to the reported decreased inflammasome activation in macrophages,
247 which is thought to contribute to the stealth phenotype of *S. Typhimurium* ST313 that involves evasion
248 of the host immune system during infection [25]. However, no significant differences in cell death due
249 to inflammasome activation were found between D23580 wild type and the D23580 *flhA*^{4/74} mutant ([S5](#)
250 [Fig](#)).

251 The transcriptomic data did offer an explanation for the reduced motility of D23580 on minimal
252 media. In the NonSPI2 condition, all flagellar genes were D23580-downregulated, with the exception of
253 the master regulators *flhDC* ([S5 Fig](#)). In contrast, in the InSPI2 and intra-macrophage condition, only
254 the flagellar class 2 genes (such as *flgA*) were significantly down-regulated. RfIP (YdiV) is a post-
255 transcriptional negative-regulator of the flagellar master transcriptional activator complex FlhD₄C₂ [44–
256 46]. We speculate that the downregulation of the flagellar regulon in NonSPI2 could be due to a
257 significant upregulation (3.5 fold-change) of *rfiP* in this low-nutrient environmental condition. This
258 differential expression was not seen in the InSPI2 growth condition which only differs from NonSPI2 by
259 a lower pH (5.8 versus 7.4) and a reduced level of phosphate [37].

260 We identified six genes and sRNAs that were D23580-upregulated in most growth conditions,
261 specifically *pgtE*, *nlp*, *ydiM*, *STM2475* (*SL1344_2438*), the ST64B prophage-encoded *SL1344_1966*,
262 and the sRNA STnc3750 (Fig 3C). Just four genes were D23580-downregulated in all conditions,
263 namely *pSLT043-5* (*SLP1_0062-4*), and *cysS* (Fig 3D). These findings confirmed that biologically
264 significant information can be extracted from the initial 17-condition experiment (Fig 2B and 2C) as
265 similar genes were up/down-regulated across the multiple conditions of the replicated experiment.

266 The transcriptomic data were interrogated to identify virulence-associated genes that were
267 differentially expressed between D23580 and 4/74. Coding genes and sRNAs located within the
268 *Salmonella* pathogenicity islands SPI-1, SPI-2, SPI-5, SPI-12 and SPI-16 showed differential
269 expression between D23580 and 4/74, in at least one growth condition (S6 Fig). The SPI-5-encoded
270 *sopB* gene (encoding a SPI-1 effector protein) and its associated chaperone gene (*pipC*) were
271 significantly D23580-upregulated in the InSPI2 and intra-macrophage conditions. In contrast, the SPI-
272 12-associated genes *STM2233-7* (*SL1344_2209-13*) were D23580-downregulated in the same two
273 growth conditions. Most SPI-2 genes were significantly D23580-upregulated in the ESP condition,
274 raising the possibility that the non-induced level of expression of SPI2 is higher in D23580 than 4/74.

275 The most highly differentially expressed genes in ESP (≥ 4 fold-change, $FDR \leq 0.001$) (S6 Fig),
276 included the D23580-upregulated genes required for itaconate degradation (*ripC*), *myo*-inositol
277 utilization (*reiD*), and proline uptake (*putA*). D23580-downregulated genes in the same growth condition
278 included those involved in uptake of uracil and cytosine (*uraA* and *codB*), melibiose utilization (*melAB*),
279 carbamoyl phosphate metabolism and pyrimidine biosynthesis (*carAB* and *pyrE1B*), nitrate reductase
280 (*napDF*), and sulfate metabolism (*cysPU* and *sbp*).

281 Genes that were differentially expressed between D23580 and 4/74 were also identified during
282 infection of RAW264.7 macrophages. The 16 genes that were most highly D23580-upregulated (≥ 4
283 fold-change, $FDR \leq 0.001$) included a β -glucosidase, *STM3775* (*SL1344_3740*); genes involved in
284 cysteine metabolism, *cdsH*; oxidation of L-lactate, *STM1621* (*SL1344_1551*); and the transcriptional
285 regulator *rcaA*. Genes that were D23580-downregulated during infection of macrophages were involved
286 in the secretion and import of siderophores (*iroC* and *iroD*), uptake of sialic acid (*nanM*), and maltose
287 or maltodextrin (*malEFK*), and *STM1630* (*SL1344_1560*), which encodes a hypothetical protein.

288 The key features of the transcriptional signature of D23580 included the differential expression of
289 the flagellar and associated genes, genes involved in aerobic and anaerobic metabolism, and iron-
290 uptake genes. Specifically, the aerobic respiratory pathway *cyoABCDE* was D23580-upregulated in the
291 anaerobic growth condition, and anaerobic-associated pathways *pdu*, *cbi* and *tdc* operons were
292 D23580-downregulated. Importantly, genes associated with the acquisition of iron through production
293 and uptake of siderophores were D23580-downregulated in the intra-macrophage environment. In
294 summary, the transcriptional signature of D23580 suggests that the biology of ST313 lineage 2 differs
295 from ST19 under anaerobic conditions *in vitro* and during infection of murine macrophages.

296 The challenge of data reproducibility in experimental science is widely acknowledged [47,48]. To
297 assess the robustness of our experiments, the RNA-seq-derived expression profiles that we generated
298 from five replicated conditions were compared with five relevant individual conditions. There was a high
299 level of correlation between the individual versus replicated datasets (correlation coefficients between
300 0.88 and 0.97) (S7 Fig). However, different levels of expression were seen between the individual and
301 the replicated ESP growth condition of D23580 for a small minority of genes. The main variations in
302 terms of functional gene groups involved cysteine metabolism, carbamoyl-phosphate and pyrimidine
303 biosynthesis, and nitrate metabolism. Variation in expression of the *ripCBA-lgl* operon was also
304 observed during anaerobic growth. We speculate that these alterations in gene expression reflect
305 experimental variations such as the use of different batches of media.

306 **1.5% of proteins were differentially expressed between D23580 and 4/74**

307 RNA-seq-based transcriptomic analysis does not reflect the translational and post-translational levels
308 of regulation [49]. To identify proteins that differentiated strains D23580 and 4/74, we used a proteomic
309 strategy that involved an LC-MS/MS platform, and analysed proteins from D23580 and 4/74 bacteria
310 grown in the ESP condition (S7 Table). A label-free quantification approach identified 66 differentially
311 expressed orthologous proteins (≥ 2 unique peptides, ≥ 2 fold-change, p -value < 0.05) (Fig 4), including
312 54 D23580-upregulated proteins and 12 D23580-downregulated proteins. The most highly D23580-
313 upregulated protein was PgtE, corroborating our previous study [20]. Up-regulated proteins included
314 those required for carbamoyl-phosphate and pyrimidine biosynthesis (CarAB and PyrIB), some SPI-1
315 proteins and associated effectors (PrgH, SipAB, InvG, SlrP, SopB, SopE2, SopA, SopD), RipAC
316 (itaconate degradation) and Lgl (methylglyoxal detoxification).

317 To identify genes that were differentially expressed at both the transcriptional and translational
318 levels, the quantitative proteomic data were integrated with the transcriptomic data. Eight D23580-
319 upregulated proteins (YciF, SopA, PgtE, STM2475, RipC, RibB, Nlp, and STM3775) were significantly
320 up-regulated in the transcriptomic data (≥ 2 fold-change, $FDR \leq 0.001$). Four differentially expressed
321 proteins (pSLT043, CysS, YgaD, MelA) were D23580-downregulated at the transcriptomic level (≥ 2
322 fold-change, $FDR \leq 0.001$) (Fig 5). Overall, 12 genes were differentially expressed at both the
323 transcriptional and protein levels.

324 **Evolution of *S. Typhimurium* ST313 involved the SNP-based inactivation of melibiose utilization** 325 **genes**

326 The melibiose utilization system consists of three genes: *melR*, which encodes an AraC-family
327 transcriptional regulator; *melA*, encoding the alpha-galactosidase enzyme; and *melB*. MelB is
328 responsible for the active transport of melibiose across the bacterial cell membrane. We found that the
329 *melAB* genes were D23580-downregulated at the transcriptomic level (Fig 6A). The differential
330 expression of *melA* was confirmed at the proteomic level (Fig 4A).

331 In strain D23580, the melibiose utilization genes contain three non-synonymous SNPs (4/74 →
332 D23580). Two are present in *melB* (Pro → Ser at the 398 AA, Ile → Val at the 466 AA) and one in *melR*
333 (Phe → Leu) (Fig 6B). The three SNPs were analysed in the context of a phylogeny of 258 genomes of
334 *S. Typhimurium* ST313 that included isolates from Malawi, as well as more distantly-related ST313
335 genomes from the UK [5] (S8 Table). All three SNPs were found to be monophyletic, allowing us to infer
336 the temporal order in which they arose and representing an accumulation of SNPs in melibiose
337 utilization genes over evolutionary time. The first SNP, *melB* I466V, was present in all 258 ST313 strains
338 tested and therefore arose first. The second SNP, in *melR*, was present in all ST313 lineage 2 and UK-
339 ST313 genomes, suggesting that it appeared prior to the divergence of these phylogenetic groups [5].
340 The final SNP, *melB* P398S, is present in all ST313 lineage 2 and a subset of UK-ST313 genomes,
341 consistent with this being last of the three mutations to arise (Fig 6C). ST313 strains can therefore be
342 classified into groups of strains containing one, two or three SNPs in melibiose utilization genes.

343 It has been reported that D23580 did not ferment melibiose whereas a ST313 lineage 1 isolate
344 (A130), *S. Typhimurium* SL1344 and *S. Typhi* Ty2 were able to utilize melibiose as a sole carbon source
345 [18]. MelB catalyzes the symport of melibiose with Na⁺, Li⁺, or H⁺ [51]. We confirmed that ST19 strains,

346 and strains belonging to the ST313 lineage 1, were positive for alpha-galactosidase activity. In contrast,
347 isolates representing the UK-ST313 lineage and the ST313 lineage 2 were unable to utilize melibiose.

348 To determine the biological role of the SNPs in the *melB* and *melR* genes, we employed a genetic
349 approach. Single nucleotide engineering was used to generate isogenic strains that reflect all three
350 melibiose gene SNP states for determination of the role of the SNP differences between ST313 lineage
351 2 and ST19 in the alpha-galactosidase (MelA)-mediated phenotypic defect (Fig 6D). Melibiose
352 utilization in D23580 was rescued by nucleotide exchange of the three SNP mutations (D23580 *melB*⁺
353 *melR*⁺) (Fig 6E). D23580 recovered its ability to grow with melibiose as the sole carbon source after
354 exchanging only the *melR* SNP with 4/74 (D23580 *melR*⁺). In contrast, D23580 did not grow in the same
355 medium when the exchange only involved the two *melB* SNPs (D23580 *melB*⁺). 4/74 lost its ability to
356 utilize melibiose as sole carbon source when we introduced the D23580 *melR* SNP (4/74 *melR*⁻ and
357 4/74 *melB*⁻ *melR*⁻). However, an exchange of the two nucleotides in *melB* did not eliminate the ability of
358 4/74 to grow in minimal medium with melibiose (4/74 *melB*⁻). These data correlated with the alpha-
359 galactosidase activity of the mutants, although a slight difference was observed between strains
360 D23580 *melR*⁺ (light green) and D23580 *melB*⁺ *melR*⁺ (green), and between strains 4/74 (green) and
361 4/74 *melB*⁻ (light green) (Fig 6F) suggesting an altered efficiency of melibiose utilization between the
362 two strains. To completely restore alpha-galactosidase activity in D23580, the reversion of the non-
363 synonymous SNPs in both the *melR* and *melB* genes was required. Our data suggest that the *melR*
364 SNP is critical for the loss of function of the melibiose utilization system.

365 In a chicken infection model, the *S. Typhimurium melA* transcript is more highly expressed in the
366 caecum than during *in vitro* growth [52]. In a chronic infection model, accumulation of melibiose was
367 observed in the murine gut after infection with *S. Typhimurium* [53]. The combination of the SNP-based
368 inactivation of melibiose catabolism with the conservation of the key SNPs in ST313 lineage 2 is
369 consistent with a functional role in ST313 virulence, and we are currently examining this possibility.

370 **A plasmid-encoded cysteinyl-tRNA synthetase is required for growth in D23580**

371 The dramatic down-regulation of the chromosomal *cysS* gene at both the transcriptomic (Fig 7A) and
372 proteomic levels (Fig 4A) was studied experimentally. The coding and non-coding regulatory regions of
373 the chromosomal *cysS* were identical at the DNA level in strains D23580 and 4/74. The chromosomal
374 *cysS* gene encodes a cysteinyl-tRNA synthetase, which is essential for cell growth in *S. Typhimurium*

375 and other bacteria [42,54,55]. To investigate *cysS* gene function, we consulted a transposon-insertion
376 sequencing (TIS) dataset for *S. Typhimurium* D23580 (manuscript in preparation). Genes that show the
377 absence or low numbers of transposon insertion sites are considered to be 'required' for bacterial
378 growth in a particular condition [55,56]. The data suggested that a functional chromosomal *cysS* was
379 not required for growth in rich medium (Fig 7B). We searched for *S. Typhimurium* D23580 genes that
380 encoded a cysteinyl-tRNA synthetase, and identified the pBT1-encoded gene, *pBT1-0241 (cysS^{pBT1})*,
381 which the TIS data suggested to be 'required' for growth in rich medium (Fig 7B).

382 To investigate cysteinyl-tRNA synthetase function in D23580, individual knock-out mutants were
383 constructed in the chromosomal *cysS* gene (*cysS^{chr}*), and the *cysS^{pBT1}* gene. These genes were 89%
384 identical at the amino acid level and 79% at the nucleotide level. The *cysS^{pBT1}* mutant was whole
385 genome sequenced to confirm the absence of secondary unintended mutations. The pBT1 plasmid was
386 also cured from D23580. We determined the relative fitness of the two *cysS* mutants and the pBT1-
387 cured strain. The D23580 wild type, D23580 Δ *cysS^{chr}* and D23580 Δ pBT1 mutants grew at similar rates
388 in LB, whilst the D23580 Δ *cysS^{pBT1}* mutant showed an extended lag phase (Fig 8A). The D23580
389 Δ *cysS^{pBT1}* mutant showed a more dramatic growth defect in minimal medium with glucose as the sole
390 carbon source (Fig 8B).

391 To determine whether the presence of the pBT1 plasmid was linked to the decrease in *cysS^{chr}*
392 expression, RNA from two biological replicates was isolated from the pBT1-cured strain in the ESP
393 growth condition. Differential expression analysis between this mutant and the wild-type D23580 strain
394 showed a significant increase in expression of *cysS^{chr}*, with TPM levels close to those seen in 4/74 (Fig
395 8C, S5 Table). These results suggested the pBT1 plasmid is responsible for the down-regulation of
396 *cysS^{chr}* expression in D23580.

397 The conservation of pBT1 was studied among 233 ST313 strains and compared to the presence of
398 the pSLT-BT plasmid which was found in all lineage 2 isolates (Fig 8D, S8 Table). Approximately 37%
399 of ST313 lineage 2 isolates carried the pBT1 plasmid. The pBT1 plasmid has rarely been seen
400 previously, but did show significant sequence similarity to five plasmids found in *Salmonella* strains
401 isolated from reptiles and elsewhere (98 to 99% nucleotide identity over 92 to 97% of the plasmid
402 sequence, accessions JQ418537, JQ418539, CP022141, CP022036 and CP022136, S1 Text).

403 Examples of essential bacterial genes located on plasmids are rare and this phenomenon has been
404 previously explored [59]. We conclude that the essentiality of the *cysS^{pBT1}* gene provides a novel
405 strategy for plasmid maintenance in a bacterial population.

406 **The SalComD23580 community data resource**

407 To allow scientists to gain new biological insights from analysis of this rich transcriptomic dataset, we
408 have made it available as an online resource for the visualization of similarities and differences in gene
409 expression between ST313 (D23580) and ST19 (4/74), using an intuitive heat-map-based approach
410 (bioinf.gen.tcd.ie/cgi-bin/salcom_v2.pl). To examine the transcriptional data in a genomic context, we
411 generated two strain-specific online browsers that can be accessed from the previous link: one for
412 D23580 and one for 4/74. The value of this type of online resource for the intuitive interrogation of
413 transcriptomic data has been described recently [60].

414 **Perspective**

415 To investigate the functional genomics of *S. Typhimurium* ST313, we first re-sequenced and re-
416 annotated the genome of the D23580 isolate. Our comparative genomic analysis of two *S. Typhimurium*
417 ST313 and ST19 isolates confirmed the findings of Kingsley *et al.* (2009) [19], identifying 856 SNPs and
418 indels, many instances of genome degradation and the presence of specific prophages and plasmids.
419 To discover the genetic differences that impact upon the biology of *S. Typhimurium* ST313, we used a
420 functional transcriptomic approach to show that the two *S. Typhimurium* pathovariants shared many
421 responses to environmental stress.

422 By investigating global gene expression in multiple infection-relevant growth conditions, we
423 discovered that 677 genes and sRNAs were differentially expressed between strains D23580 and 4/74.
424 A parallel proteomic approach confirmed that many of the gene expression differences led to alterations
425 at the protein level. The differential expression of 199 genes and sRNAs within macrophages allowed
426 us to predict functions of African *S. Typhimurium* ST313 that are modified during infection. The
427 comparative gene expression data were used to predict key phenotypic differences between the
428 pathovariants which are summarized in [S1 Table](#). The power of our experimental approach is
429 highlighted by our discovery of the molecular basis of the melibiose utilization defect of D23580, and a
430 novel bacterial plasmid maintenance system that relied upon a plasmid-encoded essential gene.

431 In the future, functional transcriptomics could shed light on the factors responsible for the phenotypic
432 differences that distinguish pathovariants of many bacterial pathogens.

433 **Materials and Methods**

434 **Bacterial strains**

435 The clinical isolate *Salmonella enterica* serovar Typhimurium D23580 was obtained from the Malawi-
436 Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi [19]. This strain, isolated
437 from blood of an HIV- child from Malawi, is used as a representative of the *Salmonella* sequence-type
438 ST313 after approval by the Malawian College of Medicine (COMREC ethics no. P.08/14/1614). *S.*
439 Typhimurium 4/74 was originally isolated from the bowel of a calf with salmonellosis [61] and is used
440 as a representative strain of *Salmonella* ST19. Other *Salmonella* strains referenced in this study are
441 listed in [S9 Table](#).

442 **Growth conditions**

443 All strains were routinely grown in Lennox broth (LB) containing 10 g/L tryptone, 5 g/L yeast extract and
444 5 g/L NaCl. Liquid bacterial cultures were incubated at 37°C 220 rpm for 16 h. Agar plates were
445 prepared with 1.5% Bacto Agar (BD Difco). To test the ability to grow with melibiose as the sole carbon
446 source, strains were grown in M9 minimal medium with 0.4% of melibiose. M9 minimal medium
447 consisted of 1x M9 Minimal Salts (Sigma-Aldrich), 2 mM MgSO₄, and 0.1 mM CaCl₂. Glucose was
448 added at a final concentration of 0.4% to M9 minimal medium to study growth behaviour of the *cysS*
449 mutants. Media were supplemented with antibiotics when required: kanamycin (Km) 50 µg/mL,
450 gentamicin (Gm) 20 µg/mL, tetracycline (Tc) 20 µg/mL, nalidixic acid (Nal) 50 µg/mL, and
451 chloramphenicol (Cm) 20 µg/mL.

452 Details for growing bacteria in the sixteen *in vitro* infection-relevant conditions and inside murine
453 RAW264.7 macrophages (ATCC TIB-71) have been published previously [37,39].

454 **Resequencing of *S. Typhimurium* D23580 genome**

455 For PacBio sequencing, *S. Typhimurium* D23580 was grown for 16 h in Lennox medium at 37°C 220
456 rpm. DNA was extracted using the Bioline mini kit for DNA purification (Bioline). Genomic quality was
457 assessed by electrophoresis in a 0.5% agarose gel at 30-35 V for 17-18 h. A 10 kb library was prepared
458 for DNA sequencing using three SMRT cells on a PacBio RSII (P5/C3 chemistry) at the Centre for

459 Genomic Research, University of Liverpool, UK. Illumina sequencing of *S. Typhimurium* D23580 was
460 performed by MicrobesNG, University of Birmingham, UK.

461 All the SNP and indel differences found between the chromosome and pSLT-BT sequences of the
462 D23580 strain used in this study (accession: XXXXXXXXXXXX) and the published D23580 (accession:
463 FN424405 and FN432031) were confirmed by PCR with external primers and subsequent Sanger
464 sequencing.

465 Draft sequences of the pBT2 and pBT3 plasmids were provided by Robert A. Kingsley [19], and
466 were used to design oligonucleotides for primer walking sequencing (all primer sequences are listed in
467 [S10 Table](#), Eurofins Genomics). Plasmid DNA from *S. Typhimurium* D23580 was isolated using the
468 ISOLATE II Plasmid Mini Kit (Bioline). For pBT2, the following oligonucleotides were used: Fw-pBT2-1
469 and Rv-pBT2-1, Fw-pBT2-2 and Rv-pBT2-2; and, for pBT3, the following oligonucleotides were used:
470 Fw-pBT3-3 and Rv-pBT3-3, Fw-pBT3-1 and Rv-pBT3-4, Fw-pBT3-4 and Rv-pBT3-2.

471 The resulting genome sequence was designated D23580_liv (accession: XXXXXXXXXXXX).

472 **Assembly of the *S. Typhimurium* D23580 complete genome**

473 HGAP3 [62] was used for PacBio read assembly of the D23580 chromosome, and for the large plasmids
474 pSLT-BT and pBT1. A hybrid assembly approach, Unicycler v0.4.5 [63], was used to combine the long
475 reads from PacBio and the short reads from the Illumina platform in order to assemble small plasmids
476 (not covered by PacBio due to size selection in library preparation) and to improve the large plasmids
477 assemblies.

478 **RNA isolation, cDNA library preparation and Illumina sequencing**

479 Total RNA from the sixteen *in vitro* growth conditions (EEP, MEP, LEP, ESP, LSP, 25°C, NaCl, bile
480 shock, low Fe²⁺ shock, anaerobic shock, anaerobic growth, oxygen shock, NonSPI2, InSPI2, peroxide
481 shock, and nitric oxide shock) and murine RAW264.7 macrophages was isolated using TRIzol, and
482 treated with DNase I, as described previously [37,39].

483 For RNA-seq, cDNA libraries were prepared and sequenced by Vertis Biotechnologie AG (Freising,
484 Germany). Briefly, RNA samples were fragmented with ultrasound (4 pulses of 30 sec at 4°C), treated
485 with antarctic phosphatase, and rephosphorylated with polynucleotide kinase (PNK). RNA fragments
486 were poly(A)-tailed and an RNA adapter was ligated to the 5'-phosphate of the RNA. First-strand cDNA

487 synthesis was carried out using an oligo(dT)-adapter primer and M-MLV reverse transcriptase. cDNA
488 was subsequently amplified by PCR to 20-30 ng/μL, and purified using the Agencourt AMPure XP kit
489 (Beckman Coulter Genomics). cDNA samples were pooled in equimolar amounts, size-selected to 150-
490 500 bp and sequenced on an Illumina HiSeq 2000 system (single-end 100 bp reads). Minor changes
491 were applied to different RNA-seq runs. For the third macrophage biological replicate of D23580, cDNA
492 was PCR-amplified to 10-20 ng/μL, size-selected to 200-500 bp, and samples were sequenced on an
493 Illumina HiSeq 2500 platform (1x100 bp). For RNA samples of D23580 and 4/74 grown in the four *in*
494 *vitro* growth conditions with three biological replicates, the third macrophage replicate of 4/74, and the
495 D23580 pBT1-cured strain, cDNA was PCR-amplified to 10-20 ng/μL and size-selected to 200-500 bp,
496 and cDNA libraries were single-read sequenced on an Illumina NextSeq 500 system using 75 bp read
497 length.

498 **Read processing and alignment**

499 The quality of each RNA-seq library was assessed using FastQC v0.11.5
500 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and then processed with Trimmomatic
501 v0.36 [64] to remove Illumina TruSeq adapter sequences, leading and trailing bases with a Phred quality
502 score below 20 and trim reads with an average base quality score of 20 over a 4 bp sliding window. All
503 reads less than 40 nucleotides in length after trimming were discarded from further analysis.

504 The remaining reads of each library were aligned to the corresponding genomes using Bowtie2
505 v2.2.9 [65] and alignments were filtered with Samtools v1.3.1 [66] using a MAPQ cut-off of 15. For *S.*
506 *Typhimurium* D23580, reads were aligned to the sequences of the chromosome and the pSLT-BT,
507 pBT1, pBT2, and pBT3 plasmids (accession: XXXXXXXXXXXX). For *S. Typhimurium* 4/74, reads were
508 aligned to the sequences of the published 4/74 chromosome, and the plasmids pSLT^{SL1344},
509 pCol1B9^{SL1344}, and pRSF1010^{SL1344} (accession: CP002487, HE654724, HE654725, and HE654726,
510 respectively). The RNA-seq mapping statistics are detailed in [S4 Table](#). Reads were assigned to
511 genomic features using featureCounts v1.5.1 [67].

512 The complete RNA-seq pipeline used for this study is described in [https://github.com/will-](https://github.com/will-rowe/rnaseq)
513 [rowe/rnaseq](https://github.com/will-rowe/rnaseq).

514 Two strain-specific browsers were generated for the visualization of the transcriptional data in a
515 genomic context online (bioinf.gen.tcd.ie/cgi-bin/salcom_v2.pl). The different tracks in each JBrowse
516 [50] were normalized using a published approach [68].

517 **Quantifying differences in expression with only one biological replicate**

518 Expression levels were calculated as Transcripts per Million (TPM) values [40,41], and generated for
519 coding genes and noncoding sRNAs in the chromosome, pSLT-BT and pBT1 plasmids for D23580
520 using the re-annotated D23580_liv genome ([S2 Table](#)). In 4/74, TPM values were determined for coding
521 genes and noncoding sRNAs in the chromosome, and the three plasmids pSLT^{4/74}, pCol1B9^{4/74}, and
522 pRSF1010^{4/74} [35]. Based on those values, and following previously described Materials and Methods,
523 the expression cut-off was set as TPM > 10 for genes and sRNAs [37].

524 For comparative analysis between the two *S. Typhimurium* strains D23580 and 4/74, TPM values
525 were obtained for the 4675 orthologous genes and noncoding sRNAs. These values were used to
526 calculate fold-changes between strains. Due to the availability of only one biological replicate per growth
527 condition, a conservative cut-off of ≥ 3 fold-change was used as a differential expression threshold
528 between strains.

529 **Differential gene expression analysis with three biological replicates**

530 Raw read counts from the 4674 orthologous coding genes and noncoding sRNAs for the three replicates
531 of the five conditions (ESP, anaerobic growth, NonSPI2, InSPI2, macrophage) for 4/74 and D23580
532 were uploaded into Degust ([S1 Data](http://degust.erc.monash.edu/)) (<http://degust.erc.monash.edu/>). Data were analysed using the
533 Voom/Limma approach [69,70] with an FDR of ≤ 0.001 and Log₂FC of ≥ 1 . Pairwise comparisons were
534 generated between the two strains for each specific condition. To remove genes with low counts across
535 all samples, thresholds of ≥ 10 read counts and ≥ 1 Counts Per Million (CPM) in at least the three
536 biological replicates of one sample were used [69,71].

537 For RNA-seq analysis of the D23580 Δ pBT1 strain grown in ESP, only two RNA-seq biological
538 replicates were used for differential expression analysis with the three biological replicates of D23580
539 WT.

540 **Sample processing for proteomics**

541 An LC-MS/MS (Q Exactive orbitrap) 4 h reversed phase C18 gradient was used to generate proteomic
542 data from six biological replicates of each strain, 4/74 and D23580, grown in the ESP condition in
543 Lennox broth. Pellet from bacterial cultures was resuspended in 50 mM phosphate buffer (pH 8),
544 sonicated (10 sec ON, 50 sec OFF, for 10 cycles at 30% amplitude), and supernatants were analysed
545 after centrifugation at 16,000 g for 20 min. Subsequent experimental procedures were performed at the
546 Centre for Proteome Research at the University of Liverpool, UK. In brief, 100 µg of protein were
547 digested (RapiGest, in-solution trypsin digestion) and 1 µg of digested protein was run on an LC-MS/MS
548 platform.

549 **Analysis of proteomic data**

550 A database was generated merging the amino acid sequences of the annotated genes in 4/74 [35] and
551 our re-annotated D23580 to allow homologous proteins as well as strain-specific proteins to be
552 identified. The merged database was clustered using the program Cd-hit and an identity threshold of
553 95% [72]. Clusters with a single protein, representing strain-specific proteins, were included in the
554 database with their accession ID. Clusters with more than one protein represented orthologues, and
555 only peptides common to all proteins of the cluster were included in the database. Common peptides
556 allowed label-free comparison of proteins that had a low level of sequence variation.

557 Raw data obtained from the LC-MS/MS platform were loaded into the Progenesis Q1 software
558 (Nonlinear Dynamics) for label-free quantification analysis. Differential expression analysis between the
559 two strains, 4/74 and D23580, is shown in [S7 Table](#). From those results (2013 proteins), multihit proteins
560 (peptides assigned to more than one protein in the same strain) were removed leaving a total of 2004
561 proteins. Cut-offs of ≥ 2 unique peptides *per* identified protein (1632 proteins), ≥ 2 fold-change
562 expression and p -value < 0.05 between strains (121 proteins) were used. Among the 121 proteins, 25
563 were 4/74-specific, 30 were D23580-specific, and 66 were encoded by orthologous genes between
564 strains.

565 **Alpha-galactosidase activity**

566 To assess alpha-galactosidase activity, strains were grown on Pinnacle™ *Salmonella* ABC
567 (chromogenic *Salmonella* medium, LabM). Bacteria that are able to produce alpha-galactosidase in the
568 absence of beta-galactosidase appear as green colonies on this medium due to the hydrolysis of X-

569 alpha-Gal. This enzymatic activity was correlated to the ability to grow in M9 minimal medium with
570 melibiose as the sole carbon source.

571 **Construction of scarless single-nucleotide substitution mutants**

572 Two strategies were used for single-nucleotide replacement as previously described [20]. For D23580
573 *flhA*^{4/74}, a single-strand DNA oligonucleotide recombination approach was used [73]. Briefly, the *flhA*-
574 474SNP oligonucleotide containing the SNP in 4/74 ('C') was used to replace the SNP in D23580 ('T').
575 The methodology followed the same strategy used for λ Red recombination explained below. After
576 electroporation of the ssDNA oligonucleotide into D23580 carrying the pSIM5-*tet* plasmid, screening for
577 D23580 recombinants was performed using a PCR with a stringent annealing temperature and primers
578 Fw-*flhA* and Rv-*flhA*-474SNP. The reverse primer contained the 4/74 SNP in *flhA*. The SNP mutation
579 in D23580 *flhA*^{4/74} was confirmed by Illumina whole-genome sequence (MicrobesNG, University of
580 Birmingham). Variant-calling bioinformatic analysis confirmed the intended mutation and the absence
581 of secondary nonintended mutations.

582 The second strategy for constructing scarless SNP mutants followed a previously described
583 approach based on the pEMG suicide plasmid [24,74]. Oligonucleotides *melR*-EcoRI-F and *melR*-
584 BamHI-R were used to PCR-amplify, in 4/74 and D23580, a *melR* region containing the SNP described
585 between strains. Additionally, primers *melB*-EcoRI-F and *melB*-BamHI-R were used for amplification,
586 in 4/74 and D23580, of a *melB* region containing the two SNPs described in this gene. PCR products
587 were cloned into the pEMG suicide plasmid and transformed into *E. coli* S17-1 λ *pir*. The resulting
588 recombinant plasmids were conjugated into 4/74 or D23580, depending on the strain that was used for
589 the PCR-amplification. For *S. Typhimurium* 4/74, transconjugants were selected on M9 minimal medium
590 with 0.2% glucose and Km. For *S. Typhimurium* D23580, transconjugants were selected on LB Cm Km
591 plates. As described previously [24], transconjugants were transformed with the pSW-2 plasmid to
592 promote the loss of the integrated pEMG by a second homologous recombination. The single-nucleotide
593 substitutions were confirmed by PCR amplification with external primers and sequencing. Mutants
594 D23580 *melR*⁺ and D23580 *melR*⁺*melB*⁺ were confirmed by Illumina whole-genome sequencing
595 (MicrobesNG, University of Birmingham). Variant-calling bioinformatic analysis confirmed the intended
596 mutations and the absence of secondary nonintended mutations in D23580 *melR*⁺*melB*⁺. Mutant
597 D23580 *melR*⁺ had a secondary nonintended synonymous mutation at the chromosomal location
598 436,081 in *STMMW_04211* (GCC → GCA).

599 **Construction of the $\Delta cysS^{chr}$ and $\Delta cysS^{pBT1}$ mutants in *S. Typhimurium* D23580 by λ Red**
600 **recombineering**

601 The D23580 mutants in *cysS^{chr}* (*STMMW_06051*) and *cysS^{pBT1}* (*pBT1-0241*) were constructed using
602 the λ Red recombination strategy [75]. The kanamycin resistance cassette (*aph*) of pKD4 was amplified
603 by PCR using the primer pairs NW_206/NW_207 and NW_210/NW_211, respectively. The resulting
604 PCR fragments were electroporated into D23580 carrying the recombineering plasmid pSIM5-*tet*
605 following the previously described methodology [20,76]. The $\Delta cysS^{chr}::aph$ mutation was transduced
606 into D23580 wild type using the high frequency transducing bacteriophage P22 HT 105/1 *int-201* [77]
607 as previously described [24]. The D23580 $\Delta cysS^{pBT1}::aph$ mutant was whole-genome sequenced using
608 the Illumina technology (MicrobesNG, University of Birmingham). Variant-calling bioinformatic analysis
609 confirmed the intended mutation and the absence of secondary nonintended mutations with the
610 exception of a six-nucleotide insertion in a noncoding region at the chromosomal position 2,755,248 (A
611 \rightarrow AGCAAGG). The Km resistance cassettes of the two recombinant strains, $\Delta cysS^{chr}::aph$ and D23580
612 $\Delta cysS^{pBT1}::aph$, were flipped-out using the FLP recombinase expression plasmid pCP20-TcR [22].

613 **Construction of the *S. Typhimurium* D23580 pBT1-cured strain**

614 The pBT1 plasmid was cured from D23580 using published methodology [78]. First, the *pBT1-0211*
615 gene of pBT1, encoding a putative RelE/StbE replicon stabilization toxin, was replaced by a I-SceI-*aph*
616 module by λ Red recombination. The I-SceI-*aph* module was amplified from pKD4-I-SceI [24] using
617 primers NW_163 and NW_164, and the resulting PCR fragment was electroporated into D23580
618 carrying pSIM5-*tet*. The resulting $\Delta pBT1-0211::I-SceI-aph$ mutants were selected on LB Km plates and
619 the mutation was transduced into D23580 wild type as described above. D23580 $\Delta pBT1-0211::I-SceI-$
620 *aph* was subsequently transformed with the I-SceI meganuclease producing plasmid pSW-2 [74], and
621 transformants were selected on LB Gm agar plates supplemented with 1 mM *m*-toluate, which induces
622 high expression of the I-SceI nuclease from pSW-2. The absence of pBT1 was confirmed by whole-
623 genome sequencing of the D23580 $\Delta pBT1$ strain (MicrobesNG, University of Birmingham).

624 **Growth curves in M9 melibiose, LB and M9 glucose media**

625 Overnight bacterial cultures were washed twice with PBS and resuspended in the specific growth
626 medium at an O.D._{600 nm} of 0.01. Growth curves of strains grown in M9 minimal medium supplemented
627 with melibiose or glucose were based on O.D. at 600 nm measurements every hour of samples growing

628 in a 96-well plate. Microplates were incubated at 37°C on an orbital shaker set at 500 rpm in a FLUOstar
629 Omega (BMG Labtech) plate reader. Only the values of the O.D._{600 nm} at 8 h were plotted for strains
630 grown in M9 melibiose medium. For the O.D._{600 nm} measurement of D23580, D23580Δ*cysS^{chr::frt}*,
631 D23580 Δ*cysS^{pBT1::frt}* and D23580 Δ*pBT1* grown in LB, an Infinite F Nano+ (Tecan) plate reader was
632 used with an orbital shaking of 432 rpm.

633 **Analysis of SNP conservation in the melibiose utilization operon**

634 The conservation of the two SNPs in *melB* and one SNP in *melR*, that distinguished the *S. Typhimurium*
635 strains D23580 and 4/74, was analysed in the genomes of 258 *S. Typhimurium* ST313 isolates from
636 Malawi and the United Kingdom. The A5 assembly pipeline [79] and ABACAS [80] were used when a
637 reference quality genome was not available. The PanSeq package allowed the identification of core
638 genome SNPs [81] and the concatenated SNP alignment served to obtain a maximum-likelihood
639 phylogenetic tree using PhyML [82]. BLASTn was used to identify the genotype of the melibiose SNPs
640 shown in [Fig 6C](#) in all genomes ([S8 Table](#)).

641 **Conservation of pBT1 and pSLT-BT plasmids among ST313 isolates**

642 For phylogenetic analysis of ST313 isolates, all available FASTQ data was downloaded from the ENA
643 using FASTQ dump v2.8.2 (accessions in [S8 Table](#), ENA access date: 01.02.2017). Data quality was
644 assessed using FastQC v0.11.5 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and then
645 processed with Trimmomatic v0.36 [64] to any adapter sequences, leading and trailing bases with a
646 Phred quality score below 20 and trim reads with an average base quality score of 20 over a 4 bp sliding
647 window. All reads less than 40 nucleotides in length after trimming were discarded from further analysis.

648 A multiple sequence alignment was generated by mapping isolate FASTQ data to the ST313
649 D23580 reference genome (pSLT-BT and pBT1 plasmids) (accession: XXXXXXXXXX) using Bowtie2
650 v2.2.9 [65]. Alignments were filtered (MAPQ cut-off 15) and then deduplicated, sorted and variant called
651 with Samtools v1.3.1 [66]. For each alignment, recombination was masked using Gubbins v2.2.0 [83]
652 and the variable sites were used to construct a maximum likelihood tree using RAxML [84]. Phylogenetic
653 trees were visualised using Figtree (<http://tree.bio.ed.ac.uk/software/figtree/>) and Dendroscope [85].
654 Coverage information was extracted from the alignment files using bedtools v2.26.0 [86] and visualised
655 using R. Results are shown in [Fig 8D](#) ([S8 Table](#)).

656 **Statistical analysis for phenotypic studies**

657 Unpaired two-tailed Student's t-tests were performed using GraphPad Prism 6.0 (GraphPad Software
658 Inc., La Jolla, CA, USA).

659 **Data availability**

660 The updated D23580 genome and annotation (D23580_liv) have been deposited in the European
661 Nucleotide Archive (ENA) repository (EMBL-EBI) under accession XXXXXXXXXX. The strain has been
662 deposited in the National Collection of Type Cultures (NCTC) from Public Health England under
663 accession number: pending.

664 The RNA-seq-derived transcriptomic data generated and re-analysed in this study have been
665 deposited in the Gene Expression Omnibus (GEO) database: accession number YYYYYYYYYY.

666 The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium
667 via the PRIDE [87] partner repository with the dataset identifier ZZZZZZZZZZ.

668 Resources for the visualization of the RNA-seq data in the 16 *in vitro* growth conditions and the intra-
669 macrophage environment are available online at bioinf.gen.tcd.ie/cgi-bin/salcom_v2.pl.

670

671 **Funding**

672 This work was supported by a Wellcome Trust Senior Investigator award (to J.C.D.H.) (Grant
673 106914/Z/15/Z). R.C. was supported by a EU Marie Curie International Incoming Fellowship (FP7-
674 PEOPLE-2013-IIF, Project Reference 628450). D.L.H. was supported by the Wenner-Gren Foundation,
675 Sweden. N.W. was supported by an Early Postdoc Mobility Fellowship from the Swiss National Science
676 Foundation (Project Reference P2LAP3_158684). Part of this work was supported by two awards from
677 the University of Liverpool Technology Directorate Voucher Scheme to R.C. and D.L.H. The BBSRC
678 grant BB/L024209/1 to R.R.C. provided support for MicrobesNG.

679 **Acknowledgments**

680 We are grateful to present and former members of the Hinton laboratory for helpful discussions,
681 particularly Aoife Colgan and Sathesh Sivasankaran, and Paul Loughnane for his expert technical
682 assistance. We appreciated Nicholas Feasey's insightful comments during this project. We thank
683 Margaret Hughes, and John Kenny from the Centre for Genomic Research at the University of Liverpool

684 for technical assistance for PacBio sequencing, and Alistair Darby for many helpful discussions. We
685 appreciate the assistance and discussions provided by Lynn McLean and Robert Beynon from the
686 Centre for Proteome Research at the University of Liverpool. We are grateful to Fritz Thümmeler (Vertis
687 Biotechnologie AG) for the construction of high-quality cDNA libraries. We also thank Gemma Langridge
688 and Nick Thomson for their help with improving the D23580 annotation.

689 Author contributions

690 R.C. and J.C.D.H. designed the study. R.C., D.L.H., C.K., W.Y.F., L.L.-L., X.Z., N.W., and S.E.C.
691 performed wet-lab experiments. R.C., S.V.O., P.B., R.R.C., W.P.M.R., A.V.P. analysed data. R.C. and
692 J.C.D.H. interpreted the data. J.H., D.M.M., and M.A.G. contributed to data interpretation. R.A.K.
693 provided draft sequences for pBT1, pBT2, and pBT3. K.H. created the *SalComD23580* website. R.C.
694 and J.C.D.H. wrote the manuscript.

695

696 References

- 697 1. Feasey NA, Dougan G, Kingsley RA, Heyderman RS, Gordon MA. Invasive non-typhoidal
698 salmonella disease: an emerging and neglected tropical disease in Africa. *Lancet*. 2012;379:
699 2489–2499. doi:10.1016/S0140-6736(11)61752-2
- 700 2. Ao TT, Feasey NA, Gordon MA, Keddy KH, Angulo FJ, Crump JA. Global Burden of Invasive
701 Nontyphoidal Salmonella Disease, 2010. *Emerg Infect Dis*. 2015;21: 941–949.
702 doi:10.3201/eid2106.140999
- 703 3. Uche IV, MacLennan CA, Saul A. A Systematic Review of the Incidence, Risk Factors and Case
704 Fatality Rates of Invasive Nontyphoidal Salmonella (iNTS) Disease in Africa (1966 to 2014). *PLoS*
705 *Negl Trop Dis*. 2017;11: e0005118. doi:10.1371/journal.pntd.0005118
- 706 4. Okoro CK, Kingsley RA, Connor TR, Harris SR, Parry CM, Al-Mashhadani MN, et al. Intra-
707 continental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan
708 Africa. *Nat Genet*. 2012;44: 1215–1221. doi:10.1038/ng.2423
- 709 5. Ashton PM, Owen SV, Kaindama L, Rowe WPM, Lane CR, Larkin L, et al. Public health surveillance
710 in the UK revolutionises our understanding of the invasive Salmonella Typhimurium epidemic in
711 Africa. *Genome Med*. 2017;9. doi:10.1186/s13073-017-0480-7
- 712 6. Almeida F, Seribelli AA, da Silva P, Medeiros MIC, dos Prazeres Rodrigues D, Moreira CG, et al.
713 Multilocus sequence typing of Salmonella Typhimurium reveals the presence of the highly
714 invasive ST313 in Brazil. *Infect Genet Evol*. 2017;51: 41–44. doi:10.1016/j.meegid.2017.03.009
- 715 7. Fields PI, Swanson RV, Haidaris CG, Heffron F. Mutants of Salmonella typhimurium that cannot
716 survive within the macrophage are avirulent. *Proc Natl Acad Sci*. 1986;83: 5189–5193.
717 doi:10.1073/pnas.83.14.5189

- 718 8. Langridge GC, Fookes M, Connor TR, Feltwell T, Feasey N, Parsons BN, et al. Patterns of genome
719 evolution that have accompanied host adaptation in Salmonella. *Proc Natl Acad Sci.* 2015;112:
720 863–868. doi:10.1073/pnas.1416707112
- 721 9. Tanner JR, Kingsley RA. Evolution of Salmonella within Hosts. *Trends Microbiol.* 2018;
722 doi:10.1016/j.tim.2018.06.001
- 723 10. Wheeler NE, Gardner PP, Barquist L. Machine learning identifies signatures of host adaptation
724 in the bacterial pathogen Salmonella enterica. *PLOS Genet.* 2018;14: e1007333.
725 doi:10.1371/journal.pgen.1007333
- 726 11. Nuccio S-P, Bäumlner AJ. Comparative Analysis of Salmonella Genomes Identifies a Metabolic
727 Network for Escalating Growth in the Inflamed Gut. *mBio.* 2014;5: e00929-14.
728 doi:10.1128/mBio.00929-14
- 729 12. Lokken KL, Walker GT, Tsohis RM. Disseminated infections with antibiotic-resistant non-typhoidal
730 Salmonella strains: contributions of host and pathogen factors. *Pathog Dis.* 2016;74.
731 doi:10.1093/femspd/ftw103
- 732 13. Ramachandran G, Perkins DJ, Schmidlein PJ, Tulapurkar ME, Tennant SM. Invasive Salmonella
733 Typhimurium ST313 with Naturally Attenuated Flagellin Elicits Reduced Inflammation and
734 Replicates within Macrophages. *PLoS Negl Trop Dis.* 2015;9: e3394.
735 doi:10.1371/journal.pntd.0003394
- 736 14. Goh YS, MacLennan CA. Invasive African nontyphoidal Salmonella requires high levels of
737 complement for cell-free antibody-dependent killing. *J Immunol Methods.* 2013;387: 121–129.
738 doi:10.1016/j.jim.2012.10.005
- 739 15. Okoro CK, Barquist L, Connor TR, Harris SR, Clare S, Stevens MP, et al. Signatures of Adaptation
740 in Human Invasive Salmonella Typhimurium ST313 Populations from Sub-Saharan Africa. *PLoS*
741 *Negl Trop Dis.* 2015;9. doi:10.1371/journal.pntd.0003611
- 742 16. Ramachandran G, Panda A, Higginson EE, Ateh E, Lipsky MM, Sen S, et al. Virulence of invasive
743 Salmonella Typhimurium ST313 in animal models of infection. *PLoS Negl Trop Dis.* 2017;11:
744 e0005697. doi:10.1371/journal.pntd.0005697
- 745 17. Singletary LA, Karlinsey JE, Libby SJ, Mooney JP, Lokken KL, Tsohis RM, et al. Loss of Multicellular
746 Behavior in Epidemic African Nontyphoidal Salmonella enterica Serovar Typhimurium ST313
747 Strain D23580. *mBio.* 2016;7. doi:10.1128/mBio.02265-15
- 748 18. Yang J, Barrila J, Roland KL, Kilbourne J, Ott CM, Forsyth RJ, et al. Characterization of the Invasive,
749 Multidrug Resistant Non-typhoidal Salmonella Strain D23580 in a Murine Model of Infection.
750 *PLoS Negl Trop Dis.* 2015;9: e0003839. doi:10.1371/journal.pntd.0003839
- 751 19. Kingsley RA, Msefula CL, Thomson NR, Kariuki S, Holt KE, Gordon MA, et al. Epidemic multiple
752 drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a
753 distinct genotype. *Genome Res.* 2009;19: 2279–2287. doi:10.1101/gr.091017.109
- 754 20. Hammarlöf DL, Kröger C, Owen SV, Canals R, Lacharme-Lora L, Wenner N, et al. Role of a single
755 noncoding nucleotide in the evolution of an epidemic African clade of Salmonella. *Proc Natl Acad*
756 *Sci U S A.* 2018;115: E2614–E2623. doi:10.1073/pnas.1714718115
- 757 21. Ramachandran G, Aheto K, Shirliff ME, Tennant SM. Poor biofilm-forming ability and long-term
758 survival of invasive Salmonella Typhimurium ST313. *Pathog Dis.* 2016;74.
759 doi:10.1093/femspd/ftw049
- 760 22. Kintz E, Davies MR, Hammarlöf DL, Canals R, Hinton JCD, van der Woude MW. A BTP1 prophage
761 gene present in invasive non-typhoidal Salmonella determines composition and length of the O-
762 antigen of the lipopolysaccharide. *Mol Microbiol.* 2015;96: 263–275. doi:10.1111/mmi.12933

- 763 23. Micoli F, Ravenscroft N, Cescutti P, Stefanetti G, Londero S, Rondini S, et al. Structural analysis
764 of O-polysaccharide chains extracted from different Salmonella Typhimurium strains. Carbohydr
765 Res. 2014;385: 1–8. doi:10.1016/j.carres.2013.12.003
- 766 24. Owen SV, Wenner N, Canals R, Makumi A, Hammarlöf DL, Gordon MA, et al. Characterization of
767 the Prophage Repertoire of African Salmonella Typhimurium ST313 Reveals High Levels of
768 Spontaneous Induction of Novel Phage BTP1. Front Microbiol. 2017;8.
769 doi:10.3389/fmicb.2017.00235
- 770 25. Carden S, Okoro C, Dougan G, Monack D. Non-typhoidal Salmonella Typhimurium ST313 isolates
771 that cause bacteremia in humans stimulate less inflammasome activation than ST19 isolates
772 associated with gastroenteritis. Pathog Dis. 2015;73. doi:10.1093/femspd/ftu023
- 773 26. Barrila J, Yang J, Crabbé A, Sarker SF, Liu Y, Ott CM, et al. Three-dimensional organotypic co-
774 culture model of intestinal epithelial cells and macrophages to study Salmonella enterica
775 colonization patterns. NPJ Microgravity. 2017;3. doi:10.1038/s41526-017-0011-2
- 776 27. Herrero-Fresno A, Wallrodt I, Leekitcharoenphon P, Olsen JE, Aarestrup FM, Hendriksen RS. The
777 Role of the st313-td Gene in Virulence of Salmonella Typhimurium ST313. PLOS ONE. 2014;9:
778 e84566. doi:10.1371/journal.pone.0084566
- 779 28. Carden SE, Walker GT, Honeycutt J, Lugo K, Pham T, Jacobson A, et al. Pseudogenization of the
780 Secreted Effector Gene sseI Confers Rapid Systemic Dissemination of S. Typhimurium ST313
781 within Migratory Dendritic Cells. Cell Host Microbe. 2017;21: 182–194.
782 doi:10.1016/j.chom.2017.01.009
- 783 29. Parsons BN, Humphrey S, Salisbury AM, Mikoleit J, Hinton JCD, Gordon MA, et al. Invasive Non-
784 Typhoidal Salmonella Typhimurium ST313 Are Not Host-Restricted and Have an Invasive
785 Phenotype in Experimentally Infected Chickens. PLoS Negl Trop Dis. 2013;7: e2487.
786 doi:10.1371/journal.pntd.0002487
- 787 30. Branchu P, Bawn M, Kingsley RA. Genome Variation and Molecular Epidemiology of Salmonella
788 enterica Serovar Typhimurium Pathovariants. Infect Immun. 2018;86: e00079-18.
789 doi:10.1128/IAI.00079-18
- 790 31. Lee S-J, Liang L, Juarez S, Nanton MR, Gondwe EN, Msefula CL, et al. Identification of a common
791 immune signature in murine and human systemic Salmonellosis. Proc Natl Acad Sci. 2012;109:
792 4998–5003. doi:10.1073/pnas.1111413109
- 793 32. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, et al. Complete genome
794 sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature. 2001;413: 852–856.
795 doi:10.1038/35101614
- 796 33. Wilmes-Riesenberg MR, Foster JW, Curtiss R. An altered rpoS allele contributes to the avirulence
797 of Salmonella typhimurium LT2. Infect Immun. 1997;65: 203–210.
- 798 34. Richardson EJ, Limaye B, Inamdar H, Datta A, Manjari KS, Pullinger GD, et al. Genome Sequences
799 of Salmonella enterica Serovar Typhimurium, Choleraesuis, Dublin, and Gallinarum Strains of
800 Well- Defined Virulence in Food-Producing Animals ▽. J Bacteriol. 2011;193: 3162–3163.
801 doi:10.1128/JB.00394-11
- 802 35. Kröger C, Dillon SC, Cameron ADS, Papenfort K, Sivasankaran SK, Hokamp K, et al. The
803 transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. Proc Natl
804 Acad Sci. 2012;109: E1277–E1286. doi:10.1073/pnas.1201061109
- 805 36. Hardt WD, Urlaub H, Galán JE. A substrate of the centisome 63 type III protein secretion system
806 of Salmonella typhimurium is encoded by a cryptic bacteriophage. Proc Natl Acad Sci U S A.
807 1998;95: 2574–2579.

- 808 37. Kröger C, Colgan A, Srikumar S, Händler K, Sivasankaran SK, Hammarlöf DL, et al. An Infection-
809 Relevant Transcriptomic Compendium for *Salmonella enterica* Serovar Typhimurium. *Cell Host*
810 *Microbe*. 2013;14: 683–695. doi:10.1016/j.chom.2013.11.010
- 811 38. Eriksson S, Lucchini S, Thompson A, Rhen M, Hinton JCD. Unravelling the biology of macrophage
812 infection by gene expression profiling of intracellular *Salmonella enterica*. *Mol Microbiol*. 47:
813 103–118. doi:10.1046/j.1365-2958.2003.03313.x
- 814 39. Srikumar S, Kröger C, Hébrard M, Colgan A, Owen SV, Sivasankaran SK, et al. RNA-seq Brings New
815 Insights to the Intra-Macrophage Transcriptome of *Salmonella Typhimurium*. *PLOS Pathog*.
816 2015;11: e1005262. doi:10.1371/journal.ppat.1005262
- 817 40. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM
818 measure is inconsistent among samples. *Theory Biosci*. 2012;131: 281–285.
819 doi:10.1007/s12064-012-0162-3
- 820 41. Wagner GP, Kin K, Lynch VJ. A model based criterion for gene expression calls using RNA-seq
821 data. *Theory Biosci*. 2013;132: 159–164. doi:10.1007/s12064-013-0178-3
- 822 42. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli*
823 K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*. 2006;2:
824 2006.0008. doi:10.1038/msb4100050
- 825 43. Minamino T. Hierarchical protein export mechanism of the bacterial flagellar type III protein
826 export apparatus. *FEMS Microbiol Lett*. 2018;365. doi:10.1093/femsle/fny117
- 827 44. Spöring I, Felgner S, Preuße M, Eckweiler D, Rohde M, Häussler S, et al. Regulation of Flagellum
828 Biosynthesis in Response to Cell Envelope Stress in *Salmonella enterica* Serovar Typhimurium.
829 *mBio*. 2018;9: e00736-17. doi:10.1128/mBio.00736-17
- 830 45. Takaya A, Erhardt M, Karata K, Winterberg K, Yamamoto T, Hughes KT. YdiV: a dual function
831 protein that targets FlhDC for ClpXP-dependent degradation by promoting release of DNA-
832 bound FlhDC complex. *Mol Microbiol*. 2012;83: 1268–1284. doi:10.1111/j.1365-
833 2958.2012.08007.x
- 834 46. Wada T, Morizane T, Abo T, Tominaga A, Inoue-Tanaka K, Kutsukake K. EAL domain protein YdiV
835 acts as an anti-FlhD4C2 factor responsible for nutritional control of the flagellar regulon in
836 *Salmonella enterica* Serovar Typhimurium. *J Bacteriol*. 2011;193: 1600–1611.
837 doi:10.1128/JB.01494-10
- 838 47. Casadevall A, Fang FC. Reproducible science. *Infect Immun*. 2010;78: 4972–4975.
839 doi:10.1128/IAI.00908-10
- 840 48. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread
841 and critical impact of batch effects in high-throughput data. *Nat Rev Genet*. 2010;11: 733–739.
842 doi:10.1038/nrg2825
- 843 49. Creecy JP, Conway T. Quantitative bacterial transcriptomics with RNA-seq. *Curr Opin Microbiol*.
844 2015;0: 133–140. doi:10.1016/j.mib.2014.11.011
- 845 50. Westesson O, Skinner M, Holmes I. Visualizing next-generation sequencing data with JBrowse.
846 *Brief Bioinform*. 2013;14: 172–177. doi:10.1093/bib/bbr078
- 847 51. Ethayathulla AS, Yousef MS, Amin A, Leblanc G, Kaback HR, Guan L. Structure-based mechanism
848 for Na⁺/melibiose symport by MelB. *Nat Commun*. 2014;5: 3009. doi:10.1038/ncomms4009
- 849 52. Harvey PC, Watson M, Hulme S, Jones MA, Lovell M, Berchieri A, et al. *Salmonella enterica*
850 Serovar Typhimurium Colonizing the Lumen of the Chicken Intestine Grows Slowly and
851 Upregulates a Unique Set of Virulence and Metabolism Genes. *Infect Immun*. 2011;79: 4105–
852 4121. doi:10.1128/IAI.01390-10

- 853 53. Kaiser BLD, Li J, Sanford JA, Kim Y-M, Kronewitter SR, Jones MB, et al. A Multi-Omic View of Host-
854 Pathogen-Commensal Interplay in Salmonella-Mediated Intestinal Infection. *PLOS ONE*. 2013;8:
855 e67155. doi:10.1371/journal.pone.0067155
- 856 54. Barquist L, Langridge GC, Turner DJ, Phan M-D, Turner AK, Bateman A, et al. A comparison of
857 dense transposon insertion libraries in the Salmonella serovars Typhi and Typhimurium. *Nucleic
858 Acids Res*. 2013;41: 4549–4564. doi:10.1093/nar/gkt148
- 859 55. Canals R, Xia X-Q, Fronick C, Clifton SW, Ahmer BM, Andrews-Polymenis HL, et al. High-
860 throughput comparison of gene fitness among related bacteria. *BMC Genomics*. 2012;13: 212.
861 doi:10.1186/1471-2164-13-212
- 862 56. Langridge GC, Phan M-D, Turner DJ, Perkins TT, Parts L, Haase J, et al. Simultaneous assay of
863 every Salmonella Typhi gene using one million transposon mutants. *Genome Res*. 2009;19:
864 2308–2316. doi:10.1101/gr.097097.109
- 865 57. Down TA, Piipari M, Hubbard TJP. Dalliace: interactive genome viewing on the web.
866 *Bioinformatics*. 2011;27: 889–890. doi:10.1093/bioinformatics/btr020
- 867 58. Barquist L, Mayho M, Cummins C, Cain AK, Boinett CJ, Page AJ, et al. The TraDIS toolkit:
868 sequencing and analysis for dense transposon mutant libraries. *Bioinformatics*. 2016;32: 1109–
869 1111. doi:10.1093/bioinformatics/btw022
- 870 59. Tazzyman SJ, Bonhoeffer S. Why There Are No Essential Genes on Plasmids. *Mol Biol Evol*.
871 2015;32: 3079–3088. doi:10.1093/molbev/msu293
- 872 60. Perez-Sepulveda BM, Hinton JCD. Functional Transcriptomics for Bacterial Gene Detectives.
873 *Microbiol Spectr*. 2018;6. doi:10.1128/microbiolspec.RWR-0033-2018
- 874 61. Rankin JD, Taylor RJ. The estimation of doses of Salmonella typhimurium suitable for the
875 experimental production of disease in calves. *Vet Rec*. 1966;78: 706–707.
- 876 62. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished
877 microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*. 2013;10:
878 563–569. doi:10.1038/nmeth.2474
- 879 63. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from
880 short and long sequencing reads. *PLOS Comput Biol*. 2017;13: e1005595.
881 doi:10.1371/journal.pcbi.1005595
- 882 64. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
883 *Bioinformatics*. 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170
- 884 65. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9: 357–
885 359. doi:10.1038/nmeth.1923
- 886 66. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map
887 format and SAMtools. *Bioinformatics*. 2009;25: 2078–2079. doi:10.1093/bioinformatics/btp352
- 888 67. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning
889 sequence reads to genomic features. *Bioinformatics*. 2014;30: 923–930.
890 doi:10.1093/bioinformatics/btt656
- 891 68. Conway T, Creecy JP, Maddox SM, Grissom JE, Conkle TL, Shadid TM, et al. Unprecedented High-
892 Resolution View of Bacterial Operon Architecture Revealed by RNA Sequencing. *mBio*. 2014;5:
893 e01442-14. doi:10.1128/mBio.01442-14
- 894 69. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools
895 for RNA-seq read counts. *Genome Biol*. 2014;15: R29. doi:10.1186/gb-2014-15-2-r29

- 896 70. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression
897 analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43: e47–e47.
898 doi:10.1093/nar/gkv007
- 899 71. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential
900 expression analysis of digital gene expression data. *Bioinformatics.* 2010;26: 139–140.
901 doi:10.1093/bioinformatics/btp616
- 902 72. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or
903 nucleotide sequences. *Bioinformatics.* 2006;22: 1658–1659. doi:10.1093/bioinformatics/btl158
- 904 73. Sawitzke JA, Costantino N, Li X-T, Thomason LC, Bubunenko M, Court C, et al. Probing cellular
905 processes with oligo-mediated recombination and using the knowledge gained to optimize
906 recombineering. *J Mol Biol.* 2011;407: 45–59. doi:10.1016/j.jmb.2011.01.030
- 907 74. Martínez-García E, Lorenzo V de. Engineering multiple genomic deletions in Gram-negative
908 bacteria: analysis of the multi-resistant antibiotic profile of *Pseudomonas putida* KT2440.
909 *Environ Microbiol.* 2011;13: 2702–2716. doi:10.1111/j.1462-2920.2011.02538.x
- 910 75. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12
911 using PCR products. *Proc Natl Acad Sci U S A.* 2000;97: 6640–6645.
- 912 76. Koskiniemi S, Pránting M, Gullberg E, Näsvall J, Andersson DI. Activation of cryptic
913 aminoglycoside resistance in *Salmonella enterica*. *Mol Microbiol.* 2011;80: 1464–1478.
914 doi:10.1111/j.1365-2958.2011.07657.x
- 915 77. Schmieger H. Phage *P22*-mutants with increased or
916 decreased transduction abilities. *Mol Gen Genet MGG.* 1972;119: 75–88.
917 doi:10.1007/BF00270447
- 918 78. de Moraes MH, Teplitski M. Fast and efficient three-step target-specific curing of a virulence
919 plasmid in *Salmonella enterica*. *AMB Express.* 2015;5. doi:10.1186/s13568-015-0139-y
- 920 79. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from
921 Illumina MiSeq data. *Bioinformatics.* 2015;31: 587–589. doi:10.1093/bioinformatics/btu661
- 922 80. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. ABACAS: algorithm-based automatic
923 contiguation of assembled sequences. *Bioinformatics.* 2009;25: 1968–1969.
924 doi:10.1093/bioinformatics/btp347
- 925 81. Laing C, Buchanan C, Taboada EN, Zhang Y, Kropinski A, Villegas A, et al. Pan-genome sequence
926 analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic
927 regions. *BMC Bioinformatics.* 2010;11: 461. doi:10.1186/1471-2105-11-461
- 928 82. Guindon S, Delsuc F, Dufayard J-F, Gascuel O. Estimating Maximum Likelihood Phylogenies with
929 PhyML. *Bioinformatics for DNA Sequence Analysis.* Humana Press; 2009. pp. 113–137.
930 doi:10.1007/978-1-59745-251-9_6
- 931 83. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic
932 analysis of large samples of recombinant bacterial whole genome sequences using Gubbins.
933 *Nucleic Acids Res.* 2015;43: e15–e15. doi:10.1093/nar/gku1196
- 934 84. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands
935 of taxa and mixed models. *Bioinformatics.* 2006;22: 2688–2690.
936 doi:10.1093/bioinformatics/btl446
- 937 85. Huson DH, Scornavacca C. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees
938 and Networks. *Syst Biol.* 2012;61: 1061–1067. doi:10.1093/sysbio/sys062

939 86. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
940 Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

941 87. Vizcaíno JA, Csordas A, del-Toro N, Dianas JA, Griss J, Lavidas I, et al. 2016 update of the PRIDE
942 database and its related tools. Nucleic Acids Res. 2016;44: D447–D456.
943 doi:10.1093/nar/gkv1145

944

945 **Figure legends**

946 **Fig 1. Comparative genomic analysis between *S. Typhimurium* 4/74 and D23580.** Plots were
947 obtained using the Circa software (<http://omgenomics.com/circa/>). (A) 4/74 and D23580 chromosomes;
948 (B) 4/74 pSLT and D23580 pSLT-BT plasmids. In both panels, 4/74 data are represented on the left
949 and D23580 data on the right. The four functional types of variants between D23580 and 4/74 are
950 shown on the right hand side of each panel.

951 **Fig 2. Inter-strain transcriptomic comparison of *S. Typhimurium* D23580 versus 4/74.** Expression
952 of orthologous coding genes and sRNAs was compared between strains D23580 and 4/74 during
953 growth in 17 infection-relevant conditions. The transcriptional expression value (TPM) for each coding
954 gene and sRNA in each condition in D23580 was divided by the TPM value for the same gene/sRNA
955 and condition in 4/74. Heat maps were obtained using the GeneSpring GX7.3 software (Agilent). Cluster
956 analysis was performed using data with ≥ 3 fold-change. (A) Heat map of the 1031 coding genes and
957 sRNAs that showed significant difference (≥ 3 fold-change) between the two strains in at least one
958 condition. (B) Heat map representing D23580-downregulated genes observed in all or most growth
959 conditions. (C) Heat map of the D23580-upregulated coding genes and sRNAs observed in most growth
960 conditions.

961 **Fig 3. Transcriptional signature of *S. Typhimurium* D23580.** Differential expression analysis of
962 orthologous coding genes and sRNAs between strains D23580 and 4/74 during growth in five infection-
963 relevant conditions. (A) Heat map highlighting biological relevant clusters. The CPM values of three
964 biological replicates for each coding gene and sRNA in each condition in D23580 were compared to
965 the CPM values for the same gene/sRNA and condition in 4/74. The heat map was obtained using
966 GeneSpring GX7.3 (Agilent). Cluster analysis was performed with CPM values of the 677 coding genes
967 and sRNAs that showed ≥ 2 fold-change and ≤ 0.001 of FDR in the differential expression analysis
968 generated using Degust in at least one condition. (B) Number of coding genes and sRNAs differentially
969 expressed in each of the five growth conditions, based on Degust results (≥ 2 fold-change, ≤ 0.001

970 FDR). (C) Venn diagram of the D23580-upregulated genes in the five growth conditions
971 (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). (D) Venn diagram of the D23580-downregulated
972 genes in the five growth conditions.

973 **Fig 4. Differentially expressed proteins between *S. Typhimurium* D23580 and 4/74.**

974 Representation of significant D23580-upregulated proteins (red dots) and D23580-downregulated
975 proteins (blue dots) in the ESP growth condition by Log₂[fold-change] and the chromosome location in
976 D23580 (≥ 2 unique peptides, ≥ 2 fold-change, p -value < 0.05). Grey dots refer to proteins that show
977 non-significant differences.

978 **Fig 5. Heat map of the 66 differentially expressed proteins between *S. Typhimurium* 4/74 and**

979 **D23580.** Data represent levels of expression at the proteomic level in the ESP growth condition, and in
980 two independent ESP RNA-seq datasets (one biological replicate versus three biological replicates).

981 **Fig 6. Melibiose phenotype differentiates the *S. Typhimurium* ST19 and ST313 strains.** (A)

982 Visualization of RNA-seq data with three biological replicates in the ESP and anaerobic growth
983 conditions using JBrowse [50] for the melibiose utilization operon. Scale of the mapped reads was 1 to
984 100. (B) Presence of three non-synonymous SNPs in the melibiose utilization genes (4/74 \rightarrow D23580).
985 (C) Accumulation of SNPs in the melibiose utilization genes during the evolution of ST313, in context
986 of a whole genome core SNP phylogeny. Isolate names, ST313 lineage and genotype for the three
987 SNPs are included in [S8 Table](#). (D) Alpha-galactosidase activity of representatives of ST19 and ST313
988 strains on Pinnacle *Salmonella* ABC medium (LabM), green = positive, colorless = negative. The colors
989 of the external circle correlate with the colors represented in the tree in "C". (E) Bacterial growth in
990 minimal M9 medium supplemented with 0.4% melibiose. Bars represent the mean of seven biological
991 replicates and standard deviation. Significant differences (****) indicate p -value < 0.0001 . (F) Alpha-
992 galactosidase activity of 4/74 and D23580 wild-type strains and corresponding mutants. The ability to
993 use melibiose is rescued in D23580 by exchange of the three SNP mutations.

994 **Fig 7. The pBT1 plasmid encodes the functional cysteinyl-tRNA synthetase in *S. Typhimurium***

995 **D23580.** (A) RNA-seq data for *cysS* in 4/74, and chromosomal and pBT1-plasmid-encoded *cysS* in
996 D23580 from the online JBrowse resources provided in this study. Scale of the mapped reads was 1 to
997 500. (B) Transposon library results for the *cysS^{chr}* and *cysS^{pBT1}* genes in D23580. Figures were obtained
998 using the Dalliace genome viewer [57]. Black arrows at the top represent genes. Each sample is

999 represented by three tracks. The first track contains blue and red lines that correspond to transposon
1000 insertion sites; blue means = orientation of the transposon is the same as the direction of the gene, red
1001 = opposite direction. The second track shows raw data for the Illumina sequencing reads. The third
1002 track highlights in red those genes that were considered 'required' for growth in that condition based on
1003 an insertion index. The insertion index was calculated for each gene as explained in [56,58], and genes
1004 with insertion index values < 0.05 were considered as 'required' for growth in the Lennox rich medium.
1005 The scale on the right represents sequence read coverage.

1006 **Fig 8. The pBT1-encoded *cysS* is required for optimal growth of *S. Typhimurium* D23580.** (A)
1007 Growth curves of D23580 WT, D23580 $\Delta cysS^{chr::frt}$, D23580 pBT1-cured strain, and D23580
1008 $\Delta cysS^{pBT1::frt}$ strains in LB medium, n = 8 (standard deviations are represented). (B) Growth curves in
1009 minimal M9 medium supplemented with 0.4% glucose, n = 5 (standard deviations are represented). (C)
1010 Comparison of *cysS^{chr}* expression levels (TPM values) of 4/74 (n = 3), D23580 (n = 3), and the D23580
1011 pBT1-cured strain (n = 2) in the ESP growth condition. Bars represent mean values and standard
1012 deviations. Significant differences (***) indicate *p*-value < 0.001. (D) The pBT1 plasmid is present in a
1013 subset of ST313 isolates of lineage 2 and in one isolate from lineage 1. Isolate names, ST313 lineage
1014 and coverage value for pBT1 and pSLT-BT are included in [S8 Table](#).

1015

1016 **Supporting Information**

1017 **S1 Table. Phenotypic features that distinguish *S. Typhimurium* ST313 from ST19 isolates from**
1018 **the literature.**

1019 **S1 Text. Supporting Materials and Methods.**

1020 **S1 Fig. Schematic representation of the RNA-seq-based comparative transcriptomic approach.**

1021 **S2 Table. Complete *S. Typhimurium* D23580 updated annotation with 4/74 orthologies.**

1022 **S3 Table. SNPs, MNPs, and indels between *S. Typhimurium* 4/74 and D23580.**

1023 **S2 Fig. Novel *S. Typhimurium* D23580 noncoding sRNAs.** Northern blots confirming the existence
1024 of novel sRNAs annotated in the BTP1 prophage region ([S1 Text](#)). For every individual sRNA, a
1025 northern blot and mapped reads in the same conditions are shown. The arrowheads indicate the most
1026 prominent bands. 5S rRNA was used as a loading control. Estimated length of the sRNAs is in brackets

1027 and was based on RNA-seq data and sequence analysis. Transcription start sites (TSS) are highlighted
1028 at the bottom.

1029 **S4 Table. RNA-seq sequence reads for *S. Typhimurium* 4/74 and D23580.**

1030 **S5 Table. TPM values for *S. Typhimurium* 4/74 and D23580 from the two RNA-seq datasets.**

1031 **S3 Fig. Transcriptional response to infection-relevant stress of *S. Typhimurium* 4/74 and D23580.**

1032 (A) Percentage of expressed genes (TPM > 10) for each individual strain. (B) Number of coding genes
1033 and sRNAs differentially expressed (fold-change ≥ 3) for each of the 17 infection-relevant conditions.

1034 (C) Heat map of the cluster analysis of all orthologous coding genes and sRNAs between the two strains

1035 obtained using GeneSpring GX7.3 (Agilent). The transcriptional expression value (TPM) for each

1036 coding gene and sRNA in each condition in D23580 was divided by the TPM value for the same

1037 gene/sRNA and condition in 4/74. (D) Bubble chart for 4/74 representing up-regulated coding genes

1038 and sRNAs versus down-regulated. The following comparisons based on TPM values were obtained

1039 for each specific condition: MEP, LEP, ESP, and LSP were compared to EEP; NaCl shock, bile shock,

1040 low Fe²⁺ shock, and anaerobic shock were compared to MEP; oxygen shock was compared to

1041 anaerobic growth; peroxide shock and nitric oxide shocks were compared to InSPI2; InSPI2 was

1042 compared to NonSPI2; and macrophage was compared to ESP. (E) Bubble chart for D23580.

1043 **S4 Fig. The Tn21-like antibiotic resistance cassette is inserted in the *mig-5* operon preventing**

1044 **expression of *rlgAb*, *rlgAa*, and *pSLT043*.** Visualization of the RNA-seq data in the 17 infection-

1045 relevant conditions from the online JBrowse resources provided in this study. Red arrows represent

1046 genes that showed upregulation in 4/74 versus D23580, and blue arrows represented D23580-

1047 downregulated genes. Scale of the mapped reads was 1 to 500. The insertion of the Tn21-like element

1048 is indicated by dotted lines.

1049 **S5 Fig. Differences in expression of the flagellar regulon between *S. Typhimurium* 4/74 and**

1050 **D23580.** (A) Heat map of the flagellar regulon genes representing the relative expression of D23580

1051 versus 4/74 in five growth conditions. TPM values were obtained from the RNA-seq dataset with only

1052 one biological replicate. (B) Swimming motility assay of 4/74, D23580, and D23580 *flhA*^{4/74} (S1 Text).

1053 Bars represent the mean of 12 independent replicates and standard deviation. Significant differences

1054 indicate ****, *p*-value < 0.0001; and **, *p*-value < 0.01. (C) Lactate dehydrogenase (LDH) cytotoxicity

1055 assay using BMDM C57BL/6 macrophages (S1 Text). Bars represent the mean of 6 independent

1056 replicates and standard deviation. Groups were compared using one-way Anova and Tukey's multiple
1057 comparisons test, significant differences indicate ****, p -value < 0.0001; and **, p -value < 0.01. (D) Heat
1058 map of the flagellar regulon using the RNA-seq data with three biological replicates. Results represent
1059 the fold-change (D23580 versus 4/74) and FDR values obtained from Degust.

1060 **S6 Table. RNA-seq results for *S. Typhimurium* 4/74 and D23580 from Degust.**

1061 **S6 Fig. Virulence-associated genes differentially-expressed between *S. Typhimurium* 4/74 and**
1062 **D23580.** (A) Heat map of the *Salmonella* pathogenicity islands genes and sRNAs that show ≥ 2 fold-
1063 change and ≤ 0.001 FDR (D23580 versus 4/74), obtained using GeneSpring GX7.3 (Agilent). The CPM
1064 values of three biological replicates for each coding gene and sRNA in each condition in D23580 were
1065 compared to the CPM values for the same gene/sRNA and condition in 4/74. (B) Differential gene
1066 expression of D23580 versus 4/74 in the early stationary phase (on the left) and the macrophage (on
1067 the right) conditions. Colors refer to fold-changes of D23580 versus 4/74 from differential expression
1068 analysis using Degust, red = D23580-upregulated, blue = D23580-downregulated. The figure includes
1069 genes that are differentially expressed ≥ 4 fold-change with ≤ 0.001 FDR (red and blue font color).
1070 Purple and light blue font colors represent up-regulated or down-regulated genes, respectively, that are
1071 related to the previous functional groups of genes, but have a fold-change ≤ 4 and ≥ 2 (≤ 0.001 FDR).

1072 **S7 Fig. Reproducibility of transcriptomic experiments.** Correlation coefficient plots of $\text{Log}_2[\text{TPM}$
1073 values] for 5 infection-relevant conditions. The "different sequencing runs" plots compare the two RNA-
1074 seq datasets (one biological replicate versus three biological replicates). The "same sequencing run"
1075 plots compare two samples of the RNA-seq dataset with three biological replicates. (A) 4/74, (B)
1076 D23580.

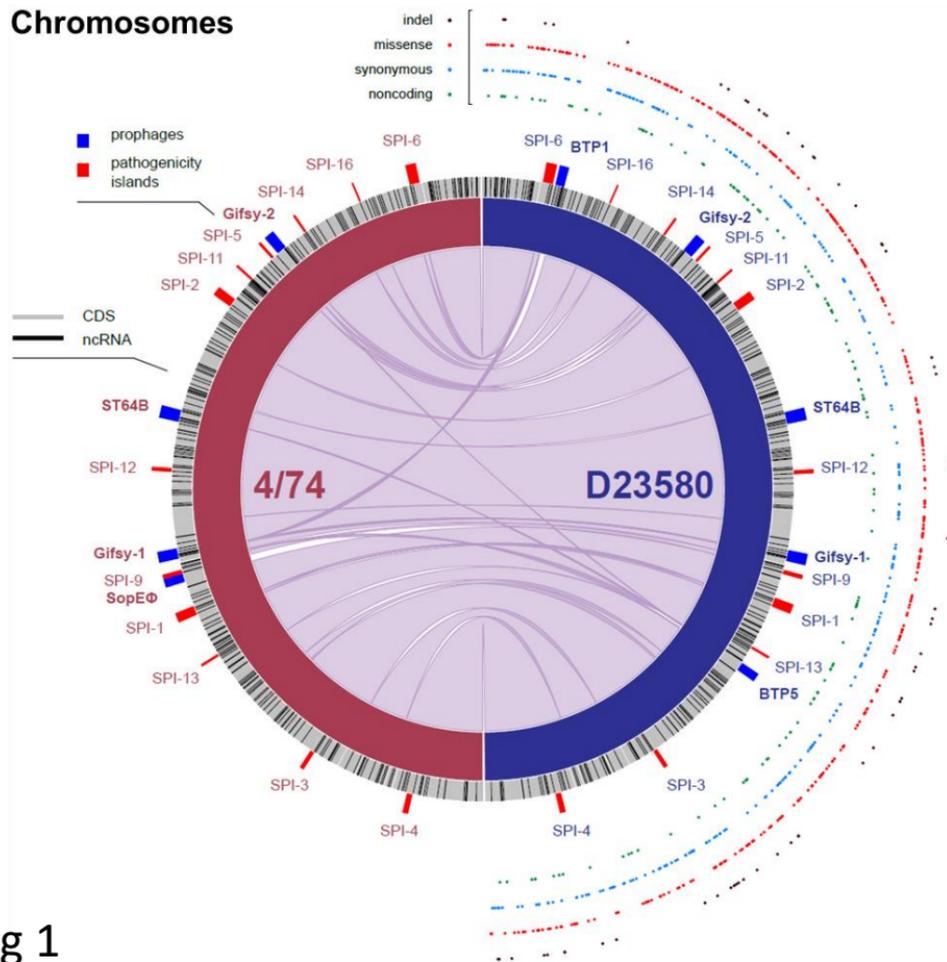
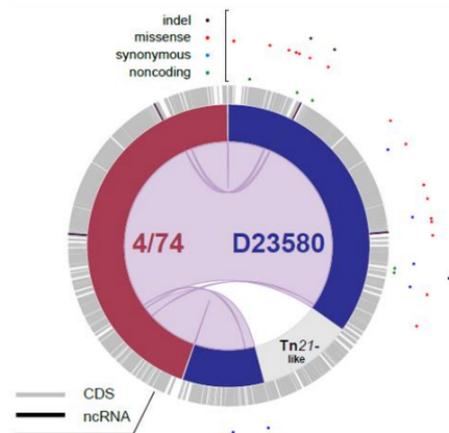
1077 **S7 Table. Proteomic data of *S. Typhimurium* 4/74 and D23580 grown in ESP.**

1078 **S8 Table. Conservation of the SNPs in the melibiose operon and the pBT1 and pSLT-BT**
1079 **plasmids among *S. Typhimurium* ST313 isolates.**

1080 **S9 Table. Bacterial strains and plasmids.**

1081 **S10 Table. Oligonucleotides used in this study.**

1082 **S1 Data. Input file for Degust.** Data include counts for the three biological replicates in five growth
1083 conditions (ESP, anaerobic growth, NonSPI2, InSPI2, macrophage) in 4/74 and D23580, and for the
1084 two biological replicates of D23580 Δ pBT1 in ESP.

A**Chromosomes****B****pSLT and pSLT-BT plasmids****Fig 1**

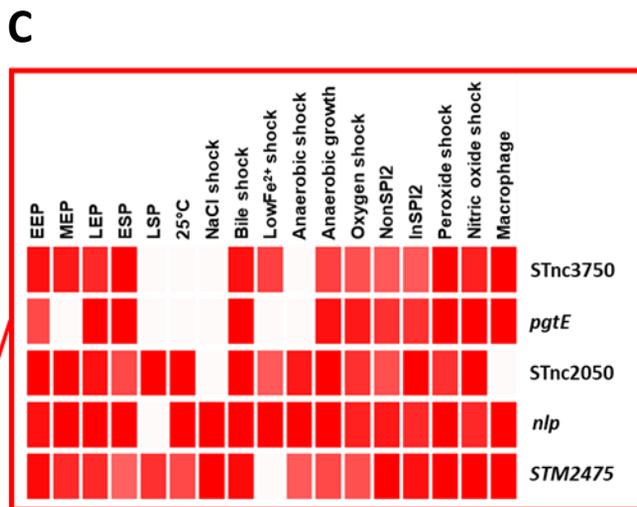
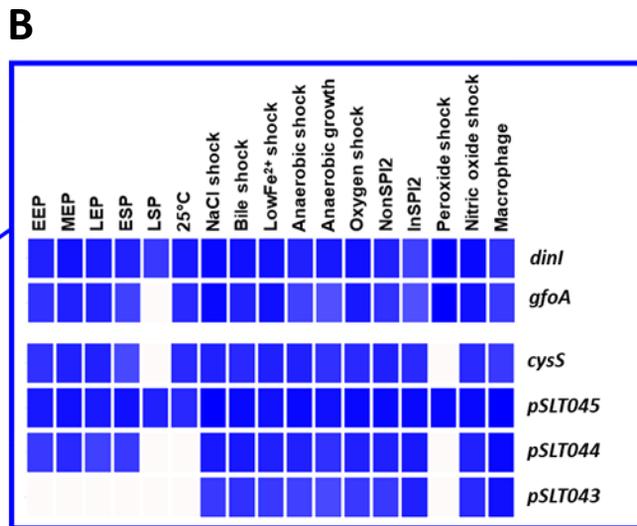
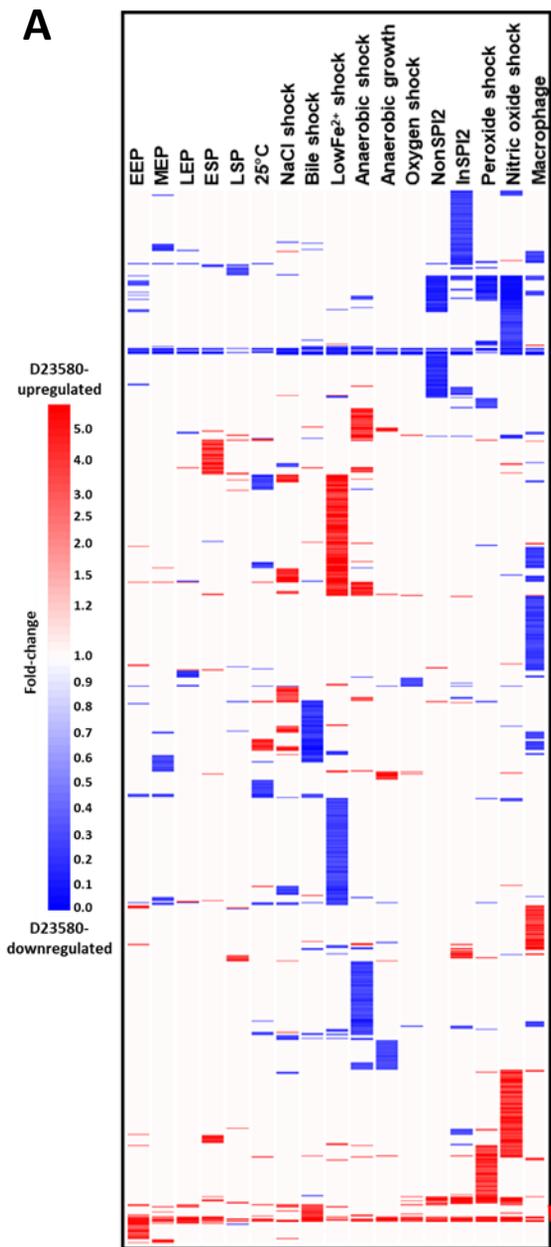
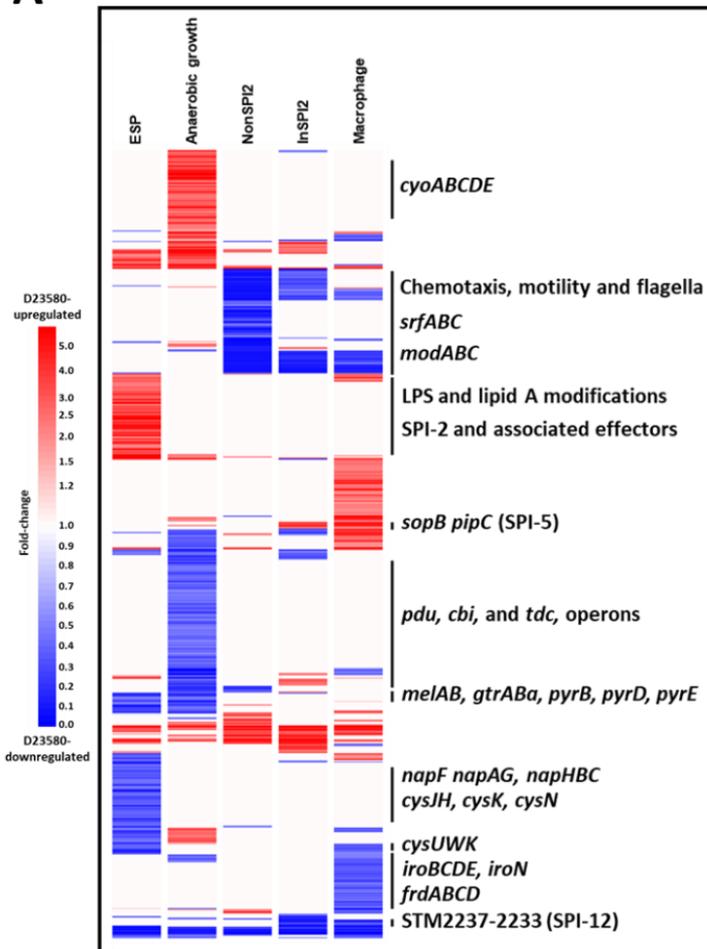
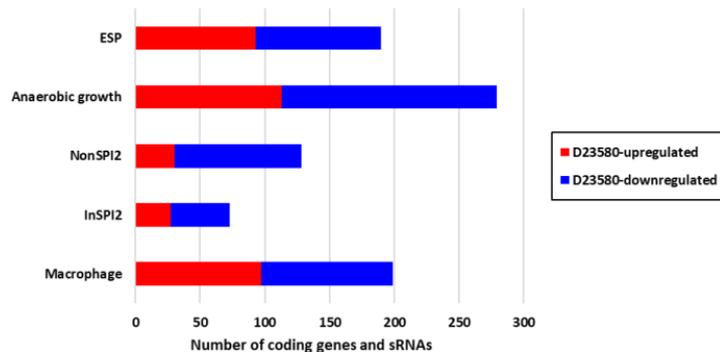


Fig 2

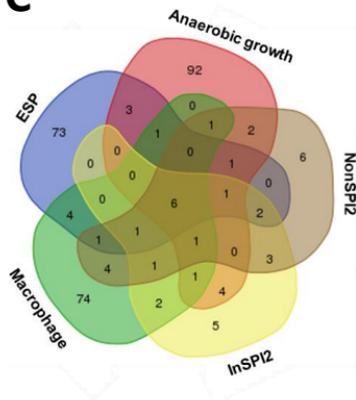
A



B



C



D

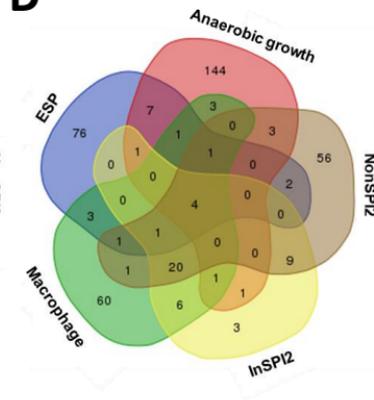


Fig 3

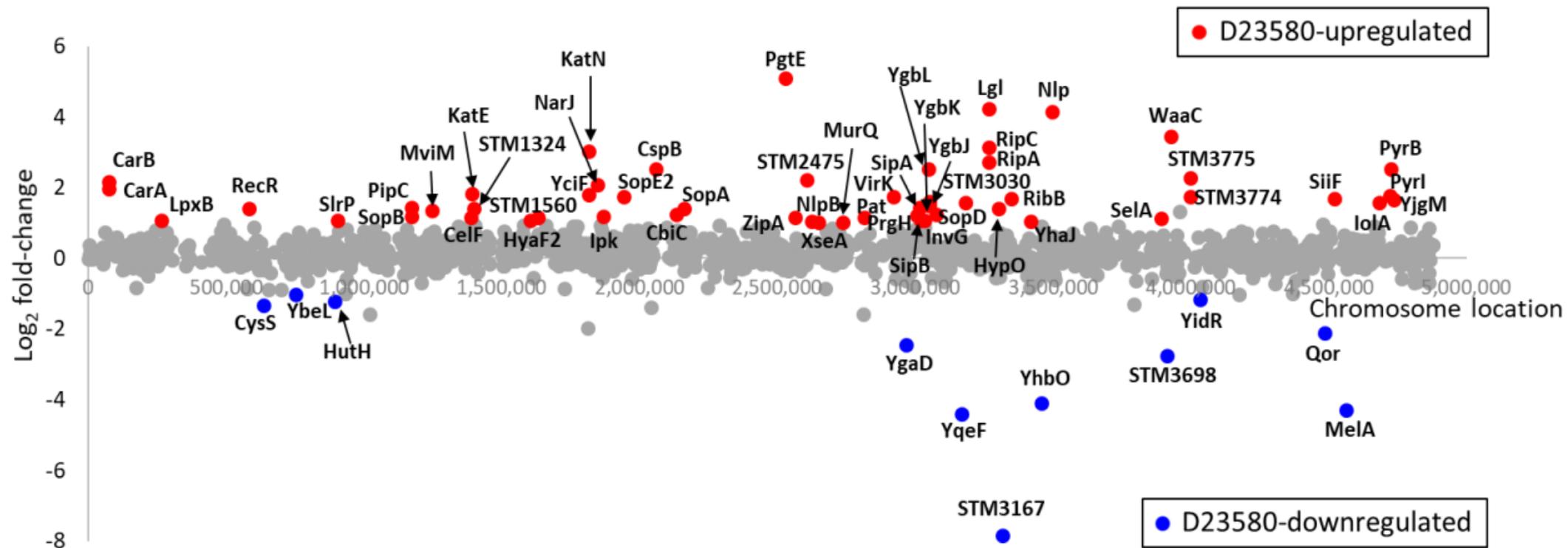


Fig 4

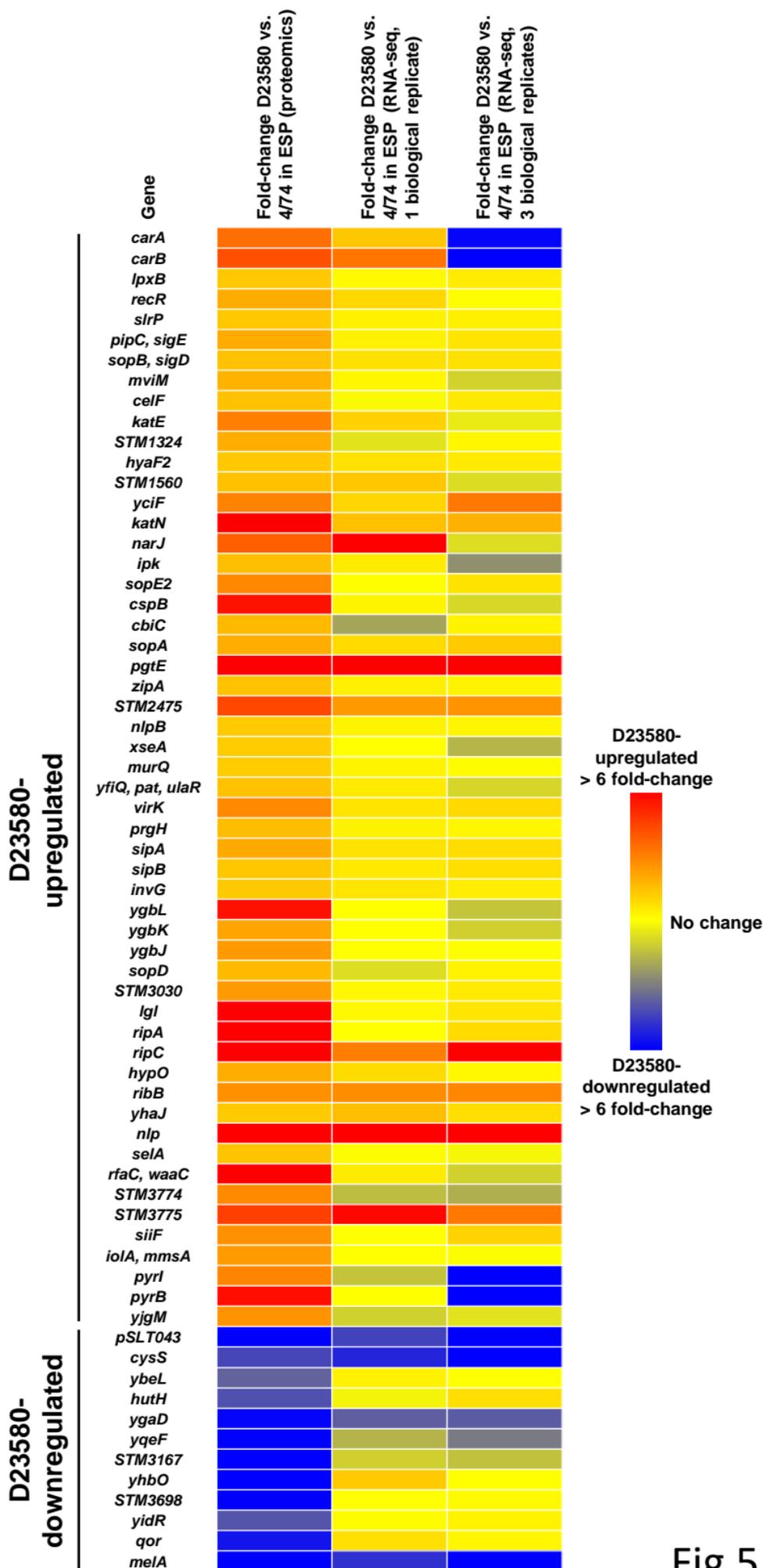


Fig 5

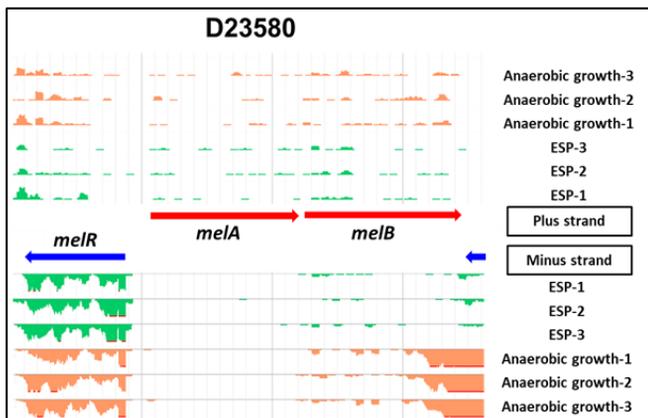
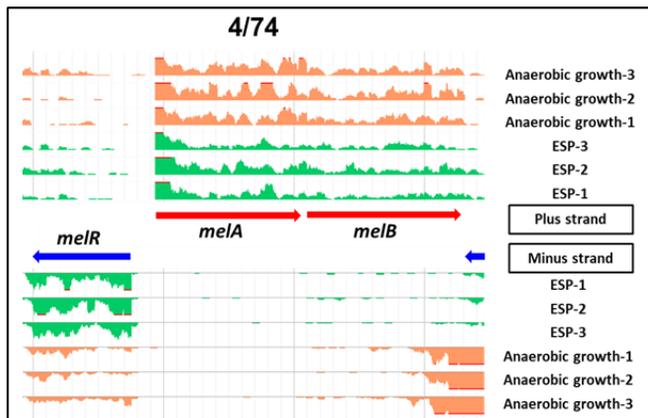
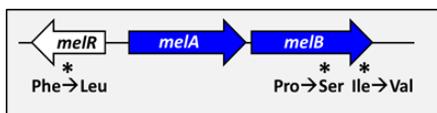
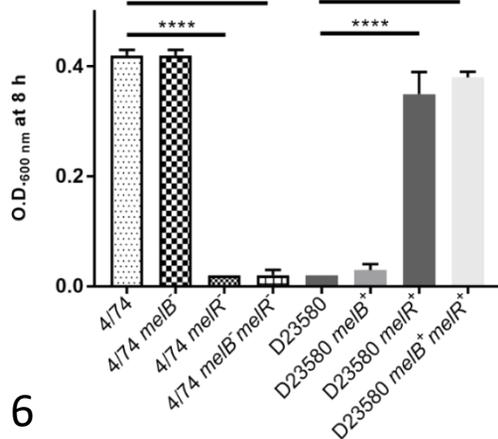
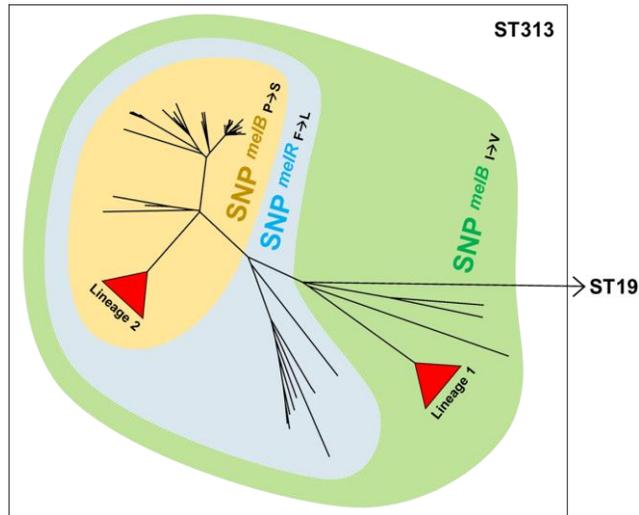
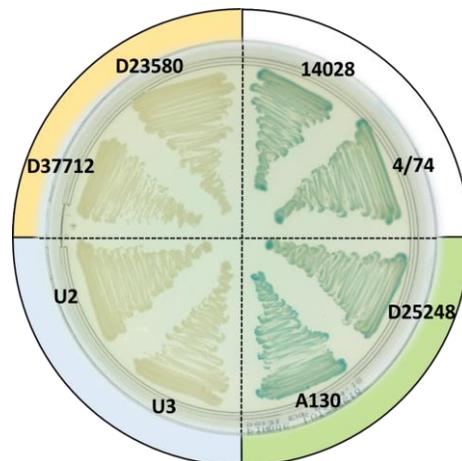
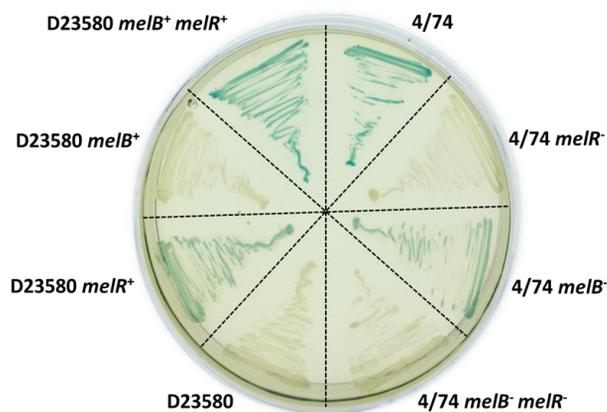
A**B****E****C****D****F**

Fig 6

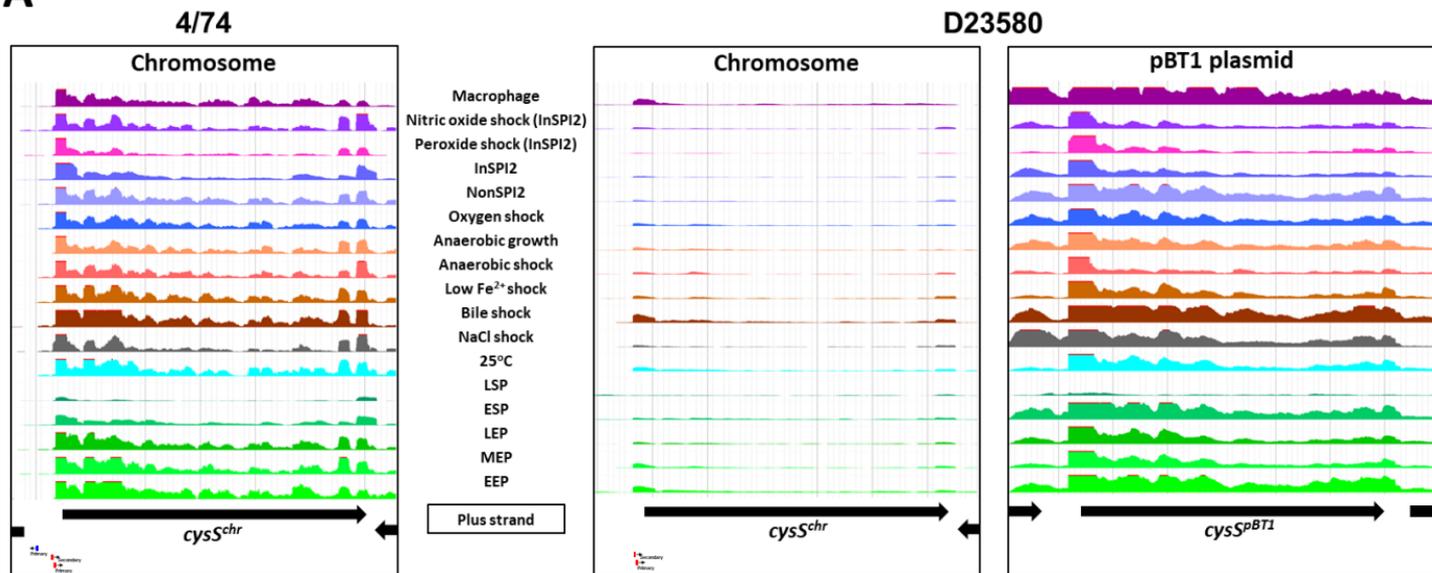
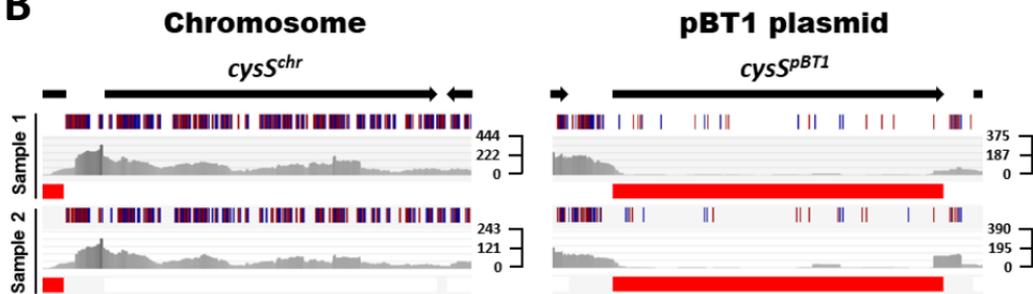
A**B**

Fig 7

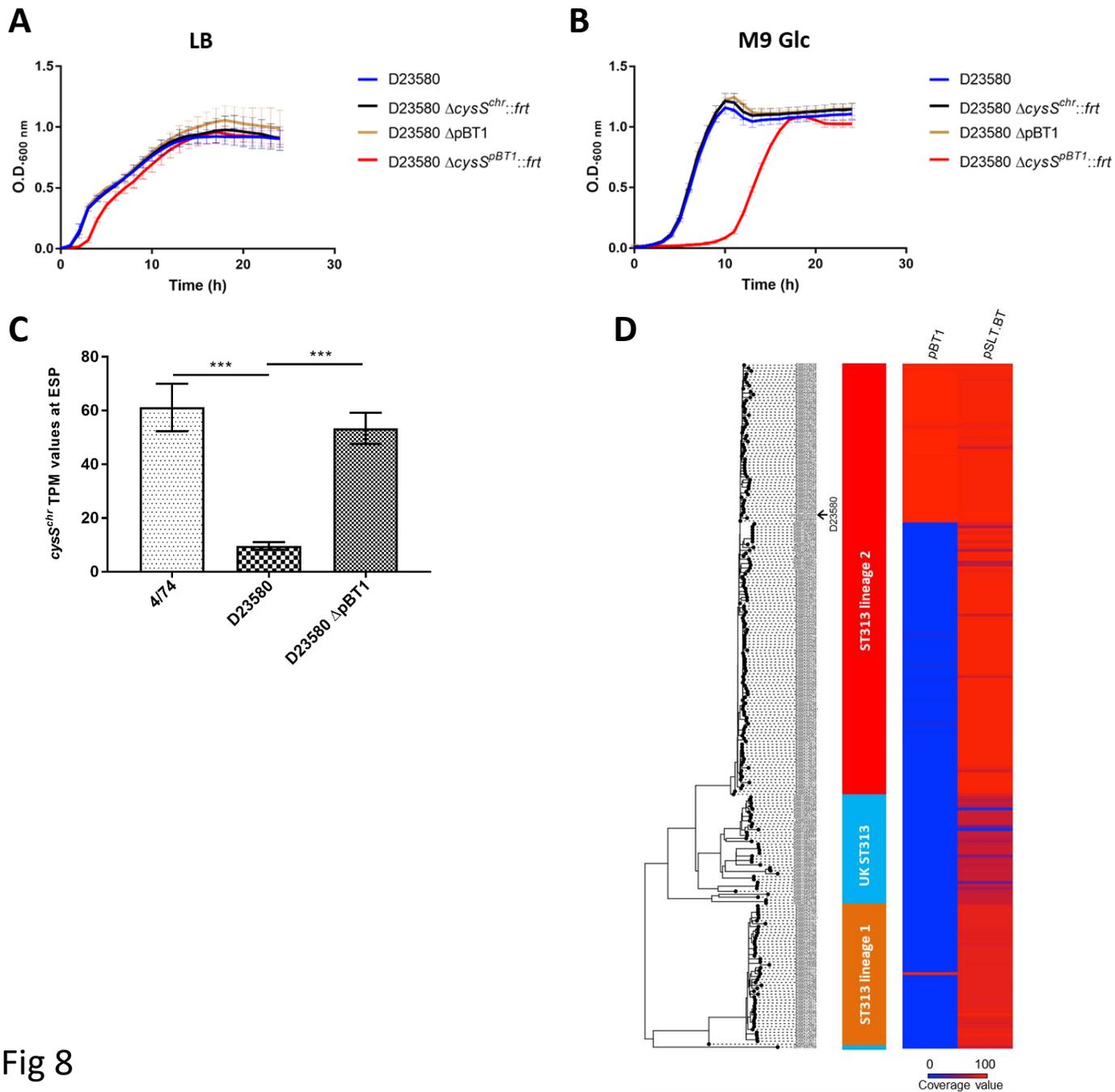


Fig 8